

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

7.9-10112
TM-79976
LACIE-C00200
VOLUME IV (REV. C)
JSC-11329

COPY 97 OF 181

(E79-10112) LARGE AREA CROP INVENTORY
EXPERIMENT (LACIE). LEVEL 3 BASELINE, CROP
ASSESSMENT SUBSYSTEM (CAS) REQUIREMENTS,
VOLUME 4, REVISION C (NASA) 145 p HC A07/MF
A01

N79-18381

Unclas
00112

"Made available under NASA sponsorship
in the interest of early and wide dis-
semination of Earth Resources Survey
program information and without liability
for any use made thereof."



NASA NOAA USDA

Level 3 Baseline

CROP ASSESSMENT SUBSYSTEM (CAS) REQUIREMENTS

NASA

National Aeronautics and
Space Administration

Lyndon B. Johnson Space Center
Houston Texas 77058

NOTICE: THIS IS A BASELINED LEVEL 3 DOCUMENT CON-
TROLLED BY THE LACIE LEVEL 3 CHANGE CONTROL BOARD.
ANY PROPOSED CHANGES SHOULD BE DOCUMENTED ON AN
RECP FORM AND TRANSMITTED TO R. B. MACDONALD,
LACIE MANAGER, NASA-JSC, CODE SF, HOUSTON, TEXAS 77058.

OCTOBER 1977

REVISIONS

REV LTR	CHANGE NO.	DESCRIPTION	DATE
		BASELINE ISSUE (Reference CCBD #III-0001, dated December 16, 1974)*	12-16-74
A	5M0052	Major revision due to rescope effort (LACIE change notices 5M0019, 5M0023, 5M0024, 5M0025, 5M0059 and 5M0065 are incorporated in this revision.)	12-16-75
B	6M0018	Major revision due to the extensive nature of LACIE change notices to this document. (LACIE change notices 6M0012, 6M0014, 6M0015, and 6M0024 are incorporated in this revision.)	7-14-76
C	7M0011	Major revision due to extensive nature of LACIE change notices to this document. (LACIE change notices 6M0018, 6M0037, and 7M0003 are incorporated in this revision.)	8-26-77

LIST OF EFFECTIVE PAGES

The current status of all pages in this document is as shown below:

<u>Page No.</u>	<u>LACIE Change Date</u>	<u>Authorizing CCBD No.</u>
<u>ii</u> thru <u>xii</u>	Revision C, 8/26/77	7M0011
<u>1-1</u> thru <u>D-28</u>	Revision C, 8/26/77	7M0011

FOREWORD

Efficient management of the Large Area Crop Inventory Experiment (LACIE) dictates that effective controls of project activities be established. To provide a basis for effective control, documentation will be prepared, baselines will be established, and changes to the baseline will be subsequently controlled by the proper management levels.

The specific control documents which will be used are defined in the LACIE Project Plan, Document LACIE-C00605. All elements of the LACIE project must adhere to these baselined control documents, and, where it is considered that the requirements should be changed, the proper change request accompanied by a full justification, must be submitted to the proper management level in accordance with established procedures. These documents will be maintained current by change notices and revisions as required. Each change notice and/or revision will reference the applicable Change Control Board Directive (CCBD) which approved the change.

This document, LACIE-C00200, Volume IV, Revision C, defines the LACIE Crop Assessment Subsystem Requirements and has been prepared in accordance with the "Instructions for Preparation of LACIE Requirements Documents," LACIE-00100, Revision C dated November 20, 1974. "Full-Up System," as used in this document, is defined as the system required to accomplish LACIE Phase II. In general, the approach used in each section is to first specify the requirements of the Full-Up System, and then to specify the requirements of any interim systems by reference to specific paragraphs in the Full-Up System requirements sections of the document. The LACIE Project Phases are defined in the LACIE Project Plan, LACIE-C00605.

The organization responsible for the implementation of each requirement defined in this document, is specified on an individual responsibility applies to the complete section, the implementation responsibility is specified after the section title. A "section" for the purposes of designating implementation responsibility is defined as being any numbered paragraph and all subparagraphs. Where different implementation responsibility apply to different portions

of a section, the implementation responsibility is specified on an individual paragraph or sentence basis, as applicable. All implementing organizations designated shall accomplish their implementation activities in accordance with the requirements specified herein.

R. B. MacDonald

for

R. B. MacDonald
Manager, Large Area Crop Inventory Experiment

CONTENTS

Section		Page
1.0	<u>FUNCTIONAL RESPONSIBILITIES</u>	1-1
1.1	<u>GENERAL</u>	1-1
1.2	<u>SPECIFIC</u>	1-2
1.2.1	<u>Sample Strategy</u>	1-2
1.2.2	<u>Wheat Area Estimation</u>	1-2
1.2.3	<u>Yield Estimates</u>	1-2
1.2.4	<u>Production Estimation</u>	1-2
1.2.5	<u>Compatibility of Crop Reporting</u>	1-3
1.2.6	<u>Accuracy Assessment</u>	1-3
2.0	<u>APPLICABLE DOCUMENTS</u>	2-1
3.0	<u>FUNCTIONAL FLOW DIAGRAMS</u>	3-1
3.1	<u>DEVELOPMENT</u>	3-1
3.1.1	<u>Full-Up System</u>	3-1
3.1.2	<u>Interim System</u>	3-1
3.2	<u>OPERATION</u>	3-1
3.2.1	<u>Full-Up System</u>	3-1
3.2.2	<u>Interim System</u>	3-1
4.0	<u>REQUIREMENTS (CONSOLIDATED)</u>	4-1
4.1	<u>CROP ASSESSMENT</u>	4-1
4.1.1	<u>Full-Up System</u>	4-1
4.1.1.1	<u>Development</u>	4-1
4.1.1.2	<u>Operations</u>	4-4
4.1.2	<u>Interim System</u>	4-10
4.2	<u>ACCURACY ASSESSMENT</u>	4-10
4.2.1.1	<u>Approach</u>	4-10
4.2.1.2	<u>System development</u>	4-10
4.2.1.3	<u>Design philosophy</u>	4-12
4.2.1.4	<u>CAS accuracy assessment operation</u>	4-13
4.2.2	<u>Interim System</u>	4-13
5.0	<u>INPUT REQUIREMENTS</u>	5-1
5.1	<u>CROP ASSESSMENT</u>	5-1
5.1.1	<u>Full-Up System</u>	5-1
5.1.1.1	<u>DAPTS</u>	5-1
5.1.1.2	<u>CAMS</u>	5-4
5.1.1.3	<u>YES</u>	5-11
5.1.1.4	<u>CAS</u>	5-13
5.1.1.5	<u>ISRRS</u>	5-15
5.1.1.6	<u>SPE</u>	5-16

Section		Page
5.1.1.7	IE	5-16
5.1.1.8	Research	5-16
5.1.1.9	Test and evaluation	5-16
5.1.2	Interim Systems	5-16
5.2	<u>ACCURACY ASSESSMENT</u>	5-16
5.2.1	<u>Full-Up System</u>	5-16
5.2.1.1	DAPTS	5-16
5.2.1.2	CAMS	5-19
5.2.1.3	YES	5-20
5.2.1.4	CAS	5-21
5.2.1.5	ISRRS	5-22
5.2.1.6	SPE	5-22
5.2.1.7	IE	5-22
5.2.1.8	Research	5-22
5.2.1.9	Test and evaluation	5-22
5.2.2	<u>Interim System</u>	5-23
6.0	<u>OUTPUT PRODUCTS</u>	6-1
6.1	<u>CROP ASSESSMENT</u>	6-1
6.1.1	<u>Full-Up System</u>	6-1
6.1.1.1	DAPTS	6-1
6.1.1.2	CAMS	6-2
6.1.1.3	YES	6-3
6.1.1.4	CAS	6-3
6.1.1.5	ISRRS	6-6
6.1.1.6	SPE	6-6
6.1.1.7	IE	6-6
6.1.1.8	Research	6-6
6.1.1.9	Test and evaluation	6-6
6.1.2	<u>Interim System</u>	6-7
6.2	<u>ACCURACY ASSESSMENT</u>	6-7
6.2.1	<u>Full-Up System</u>	6-7
6.2.1.1	Accuracy estimates	6-7
6.2.1.2	Discrepancy evaluation	6-7
6.2.1.3	System testing	6-7
6.2.2	<u>Interim Systems</u>	6-7
7.0	<u>INTERFACE REQUIREMENTS</u>	7-1
8.0	<u>OPERATIONAL REQUIREMENTS</u> <u>AFFECTING THE DESIGN</u>	8-1
8.1	<u>CROP ASSESSMENT</u>	8-1
8.1.1	<u>Throughput Requirements</u>	8-1
8.1.2	<u>Response Requirements</u>	8-1
8.1.3	<u>Reliability Requirements</u>	8-1
8.1.3.1	Data base	8-1

Section		Page
8.1.4	<u>Security Requirements</u>	8-1
8.1.5	<u>Delivery Requirements</u>	8-2
8.1.6	<u>Quality Assurance Requirements</u>	8-2
8.1.7	<u>Other Operational Requirements</u>	8-2
8.2	<u>ACCURACY ASSESSMENT</u>	8-2
8.2.1	<u>Throughput Requirements</u>	8-2
8.2.2	<u>Response Requirements</u>	8-3
8.2.3	<u>Reliability Requirements</u>	8-3
8.2.4	<u>Security Requirements</u>	8-3
8.2.5	<u>Delivery Requirements</u>	8-3
8.2.6	<u>Quality Assurance Requirements</u>	8-3
9.0	<u>VERIFICATION REQUIREMENTS</u>	9-1
9.1	<u>CROP ASSESSMENT</u>	9-1
9.2	<u>ACCURACY ASSESSMENT</u>	9-1
10.0	<u>RESEARCH REQUIREMENTS</u>	10-1
10.1	<u>CROP ASSESSMENT</u>	10-1
10.1.1	<u>Pre-Phase I</u>	10-1
10.1.2	<u>Desirable Approaches</u>	10-1
10.1.2.1	<u>Sampling</u>	10-1
10.1.2.2	<u>Aggregation</u>	10-2
10.1.2.3	<u>Harvested Wheat</u>	10-3
10.2	<u>ACCURACY ASSESSMENT</u>	10-3
10.2.1	<u>Error Model</u>	10-4
11.0	<u>TEST AND EVALUATION REQUIREMENTS</u>	11-1
11.1	<u>CROP ASSESSMENT</u>	11-1
11.2	<u>ACCURACY ASSESSMENT</u>	11-1
Appendix		
A	<u>SAMPLE SEGMENT ALLOCATION</u> <u>PROCEDURE FOR U.S.A.</u>	A-1
B	<u>SAMPLING AND ESTIMATION PLAN</u>	B-1
C	<u>HARTLEY-RAO PROCEDURE FOR PPS</u> <u>SAMPLING</u>	C-1
D	<u>SAMPLING AND ESTIMATION PROCEDURE BASED</u> <u>ON NATURAL STRATA</u>	D-1

FIGURES

Figure		Page
3-1	CAS subsystem development of crop assessment	3-2
3-2	CAS accuracy assessment full-up system development functional flow diagram	3-3
3-3	Functional flow diagram of electronic crop area estimate within CAS	3-4
3-4	Functional flow diagram of electronic yield estimate within CAS	3-5
3-5	Functional flow diagram of electronic production estimate within CAS	3-6
3-6	CAS accuracy assessment full-up system operational functional flow diagram	3-7

GLOSSARY

ADP	Automatic Data Processing
AES	Application Evaluation System
AI	Analyst Interpreter
A&M	Agricultural and Mechanical
ASCS	Agriculture Stabilization and Conservation Service
ASVB	Applications System Verification Branch
CAS	Crop Assessment Subsystem
CAMS	Classification and Mensuration Subsystem
CATEGORY 1	Category 1 requirements are those which can be understood to the degree that definite action can be taken to satisfy the requirement; i.e., the requirement is defined to the extent required for implementation by the implementing organization.
CATEGORY 2	Category 2 requirements are those which can be understood to the degree that action can be initiated but further definition of the requirements will be necessary for complete implementation.
CATEGORY 3	Category 3 requirements define a general area in which a requirement exists but the requirement needs to be further defined before implementation can be initiated.
CATEGORY 4	Category 4 requirements are those which may or may not be levied at some time in the future; i.e., they consist of items which are still being considered for possible implementation.
CCBD	Change Control Board Directive
CCT	Computer Compatible Tape
CITARS	Crop Identification Technology Assessment for Remote Sensing

CPU	Central Processing Unit
CRD	Crop Reporting District
CS	Civil Servant
DAPTS	Data Acquisition, Preprocessing and Transmission Subsystem
DSAD	Data Systems and Analysis Directorate
E	East
EOD	Earth Observations Division
EP	Error Parameters
ERS	Equipment Record System
FAS	Foreign Agriculture Service
Full-Up System	System required to accomplish LACIE Phase II
FY	Fiscal Year
GDSD	Ground Data Support Division
GSFC	Goddard Space Flight Center
GSR	Green Support Room
ICD	Interface Control Document
ID	Identification
IE	Information Evaluation
Interim System	System requirements to accomplish subphases of LACIE Phase I, excluding subphases IA and IB
I/O	Input/Output
ISRRS	Information Storage, Retrieval, and Reformatting Subsystem
JSC	Johnson Space Center

K	Thousands (kilo)
LACIE	Large Area Crop Inventory Experiment
LACIE PO	LACIE Project Office
Landsat	Land Satellite
LEC	Lockheed Electronics Company
MSS	Multispectral Scanner
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
Out of Sync	Not synchronized
ONC	Operational Navigational Charts
PFC	Production Film Converter
PPS	Probability Proportional to Size
R&D	Research and Development
Pseudo- county	Aggregates of 5- by 6-mm rectangles whose center points lie within actual county boundaries
RTEB	Research Test and Evaluation Branch
S	South
S&AD	Science and Applications Directorate
SPE	System Performance Evaluation
SRS	Statistical Reporting Service
Standard Statistics	Standard deviation, coefficient of variation, bias, and confidence level
T&E	Test and Evaluation
TBD	To Be Determined

USDA	United States Department of Agriculture
VIP	Very Important Person
W	West
WMO	World Meteorological Organization
YES	Yield Estimation Subsystem

SECTION 1.0

FUNCTIONAL RESPONSIBILITIES

1.1 GENERAL

The CAS will consist of multidiscipline analyst/software capabilities for which responsibilities include:

- A. Providing, on a monthly basis, the area, yield, and wheat production from planting through post-harvest. Phase I responsibilities will be area estimation for the Applications Evaluation effort, as well as yield and production for the T&E and R&D efforts. Phase II responsibilities will include area, yield, and production reports in the AES.
- B. Development of a sampling strategy and definition of sample segments for area and production calculations.
- C. Insuring that related LACIE subsystem's outputs will be properly compatible within the crop reporting system for eventual transfer into a USDA operational environment.
- D. Providing criteria and procedures for the assessment and evaluation of LACIE reports in terms of USDA requirements and in consideration of established performance goals and cost benefit criteria.
- E. Investigating discrepancies in the LACIE area, yield, and production reports identified by the USDA evaluation of the information contained in these reports. (If the source of the discrepancy is determined to be within the AES, the CAS will analyze and report the causes and make recommendations.)

1.2 SPECIFIC

1.2.1 Sample Strategy

- A. Definition and delineation of agricultural sampling areas
- B. Insure adequacy of sampling strategy for LACIE requirements
 - 1. Develop and utilize a sample strategy to select the number and location of the sample segments required to estimate accurately the wheat area.
 - 2. The sample strategy will be tested on a limited basis prior to Phase I.
 - 3. The results of the sample strategy will be monitored during Phase I operations and the strategy will be improved as required for use in Phase II.

1.2.2 Wheat Area Estimation

In Phases I and II, the CAS will incorporate CAMS wheat proportion estimates into the crop area aggregation model to obtain monthly estimates of area by strata, zone, region, and country.

1.2.3 Yield Estimates

In Phase II, the CAS will utilize the stratum and zone average yields supplied by the YES to provide monthly estimates of the yield for the regions and countries.

1.2.4 Production Estimation

In Phase II, the CAS will operate the crop-production model to obtain monthly estimates of crop production based on the products of area and yield determined by the CAMS and the YES, aggregated to production values for strata, zones, regions, and country. Production estimates for a geographical area (country,

region, etc.) will be an upward aggregation of production estimates of its subareas.

1.2.5 Compatibility of Crop Reporting

The CAS will insure that the LACIE output reports are compatible with the USDA crop reporting system to allow eventual transfer into the USDA operation. The CAS will provide the criteria and procedures for the assessment, and evaluation of the LACIE output products to meet USDA requirements.

1.2.6 Accuracy Assessment

The CAS will monitor and evaluate the development activities and products of the CAMS, CAS, and YES, and track AES parameters in sufficient detail to provide an assessment of the reliability and accuracy of the area, yield, and production estimates issued by the LACIE. Any deficiencies will be reported to the AES and Quality Assurance Managers.

The CAS will define the ground truth requirements necessary to support this activity. The CAS will also provide assistance to the Washington USDA LACIE staff in the following areas:

- A. Identify additional information processing support software/hardware requirements for the USDA based on operating experience and user requirements.
- B. Develop T&E criteria and procedures for the LACIE system.
- C. Supervise the application of criteria and procedures in B. to a USDA test bed system.
- D. Participate in and monitor ongoing research and development efforts.

SECTION 2.0

APPLICABLE DOCUMENTS

The following documents are applicable to the extent specified herein:

- A. LACIE Project Plan, LACIE-C00605, dated August, 1975.
- B. Instruction for Preparation of LACIE Requirements Documents, LACIE-00100, Revision C, dated November 20, 1974.
- C. GSFC/JSC Interface Control Document, LACIE-00701, dated March 20, 1975.
- D. LACIE Requirements Documents, LACIE-C00200 (Revision A) Volumes IA, IB, IC, ID, IE, II, IV, V, VI-A, dated December, 1975 and LACIE-00200 Volumes III, VI-B, VII, VIII, IX, dated December, 1974.
- E. LACIE Subsystem Procedures Documents - TBD.
- F. LACIE Operations Plan, LACIE-C00602, dated July 11, 1975 and LACIE Operations Plan, Phase II, LACIE-C00606, dated September, 1975.
- G. LACIE Rescope Document, LACIE-00401, Volume I, dated April, 1975.
- H. USDA/NASA Interface Control Documents, Volume I, LACIE-C00705, dated April, 1976, and Volume II (Revision A), LACIE-C00707, dated April, 1976.
- I. LACIE User's Manual.
- J. NOAA/NASA/USDA Interface Control Document, LACIE-C00710, dated December, 1975.
- K. CAS/IE Interface Control Document, LACIE-C00709, dated December, 1975.

PRECEDING PAGE BLANK

SECTION 3.0

FUNCTIONAL FLOW DIAGRAMS

3.1 DEVELOPMENT

3.1.1 Full-Up System

The development of the CAS crop assessment operational functions is shown in figure 3-1, and the development of the accuracy assessment is shown in figure 3-2.

3.1.2 Interim System

The development of the Phase IC portion of CAS crop assessment is shown in figure 3-1 as the top two flow diagrams for sample strategy and area calculation. No operational accuracy assessment is being developed for the interim system.

3.2 OPERATION

3.2.1 Full-Up System

The full-up operational CAS crop assessment is shown in figure 3-3 to 3-5. The operational accuracy assessment flow is shown in figure 3-6.

3.2.2 Interim System

The Phase IC portion of CAS crop assessment is shown in figure 3-3 (area only). No operational accuracy assessment is conducted on the interim system.

CAS SUBSYSTEM DEVELOPMENT

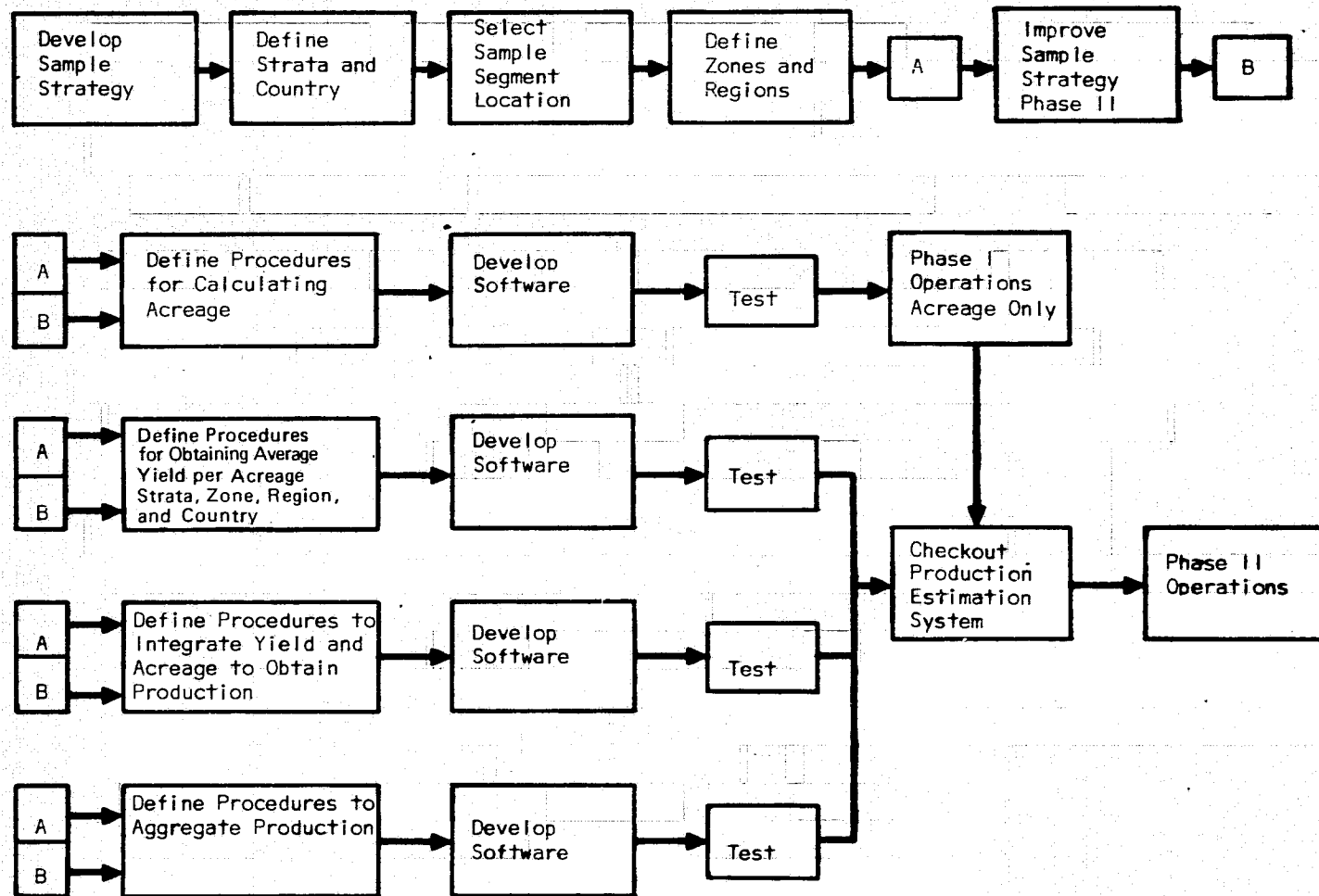


Figure 3-1.- CAS subsystem development of crop assessment.

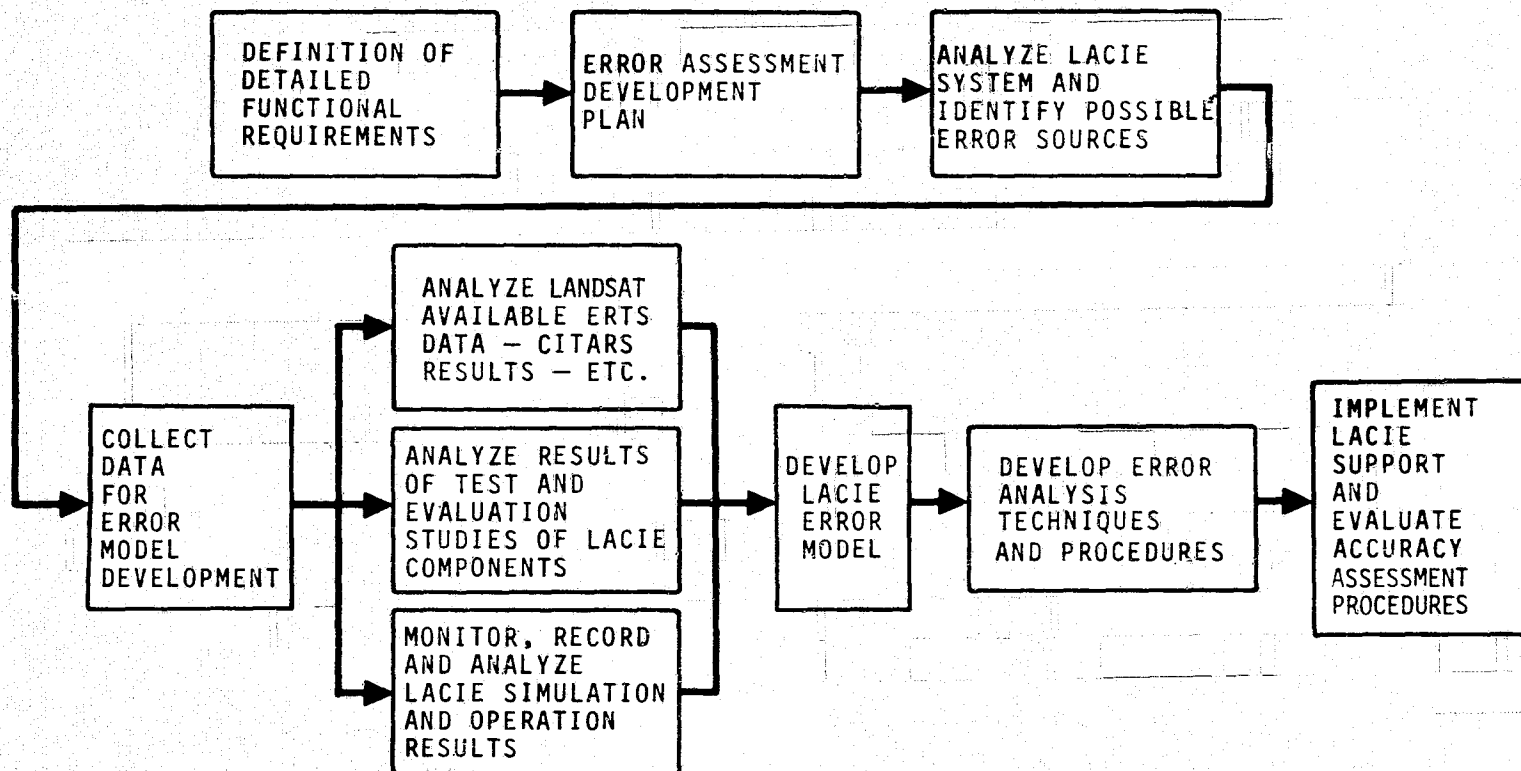


Figure 3-2.- CAS accuracy assessment full-up system development functional flow diagram.

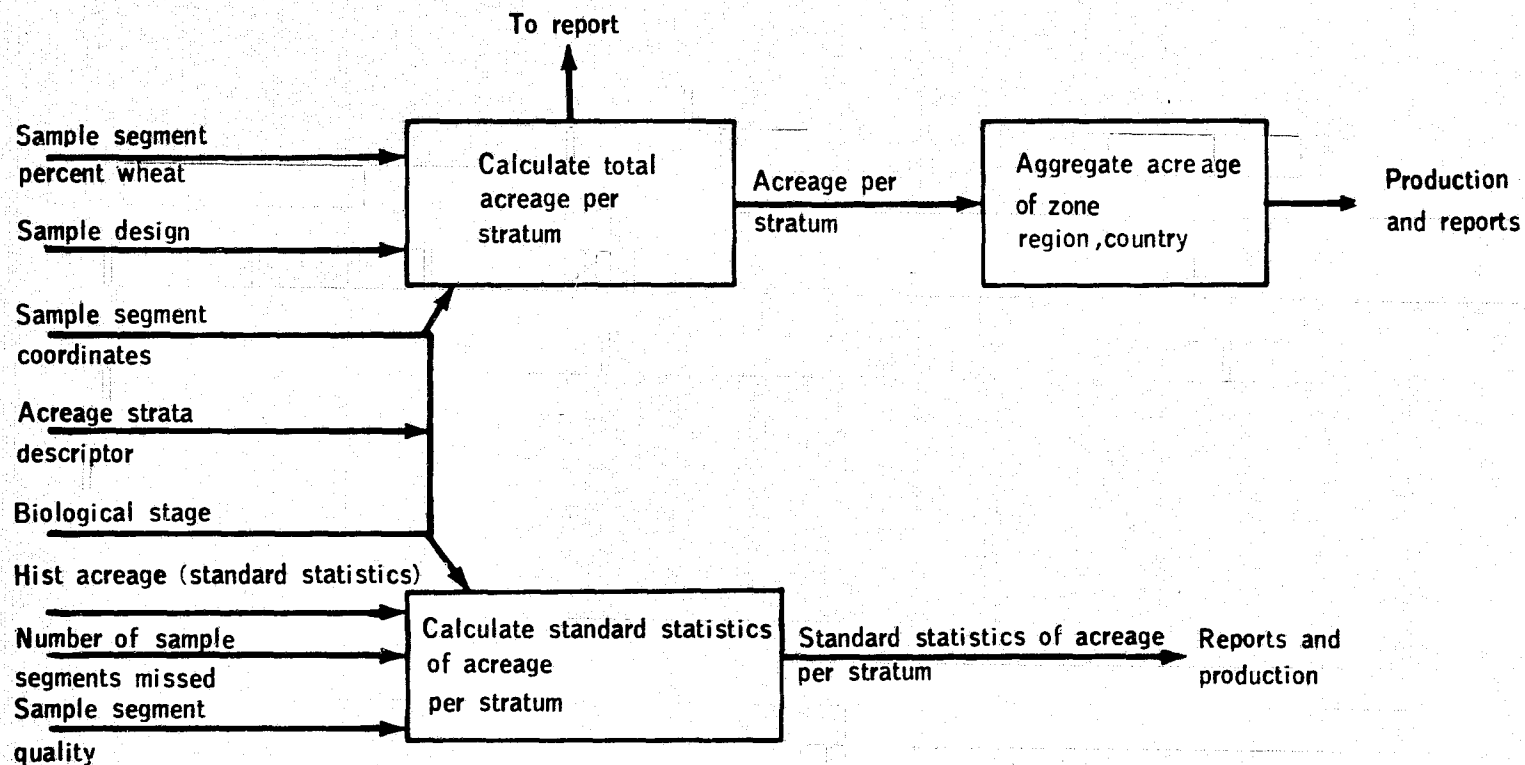


Figure 3-3.- Functional flow diagram of electronic crop area estimate within CAS.

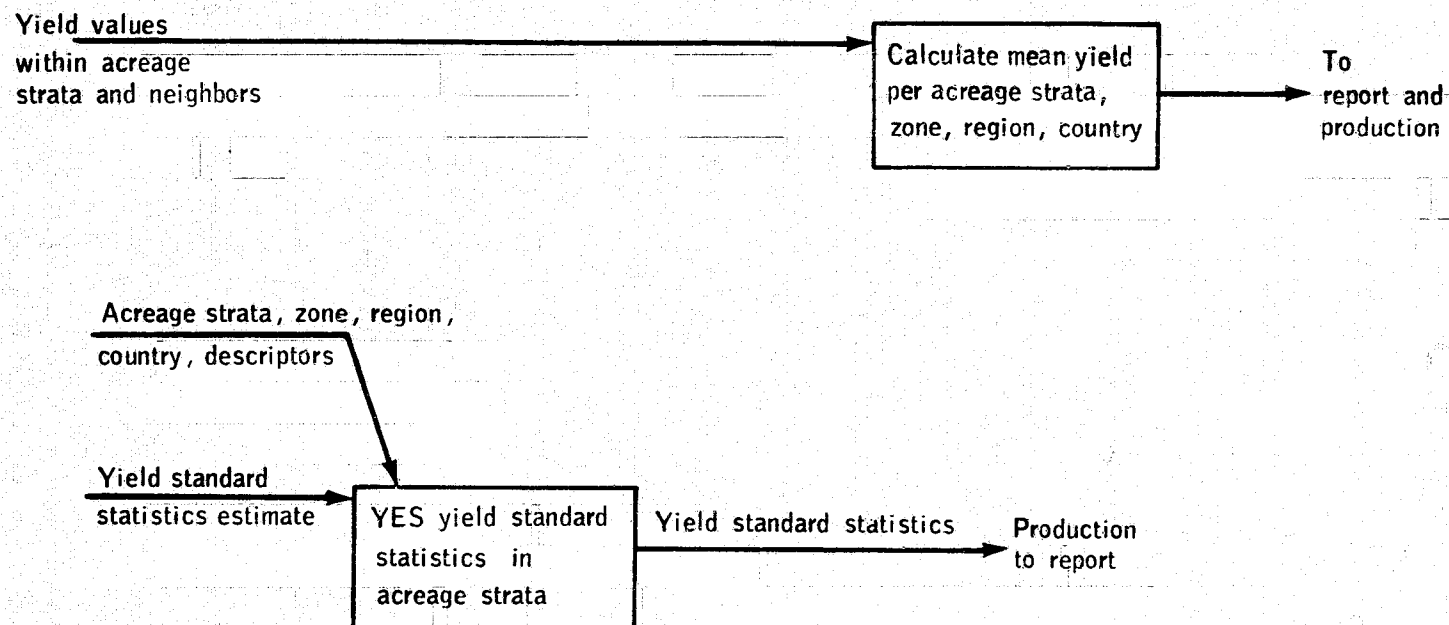


Figure 3-4.- Functional flow diagram of electronic yield estimate within CAS.

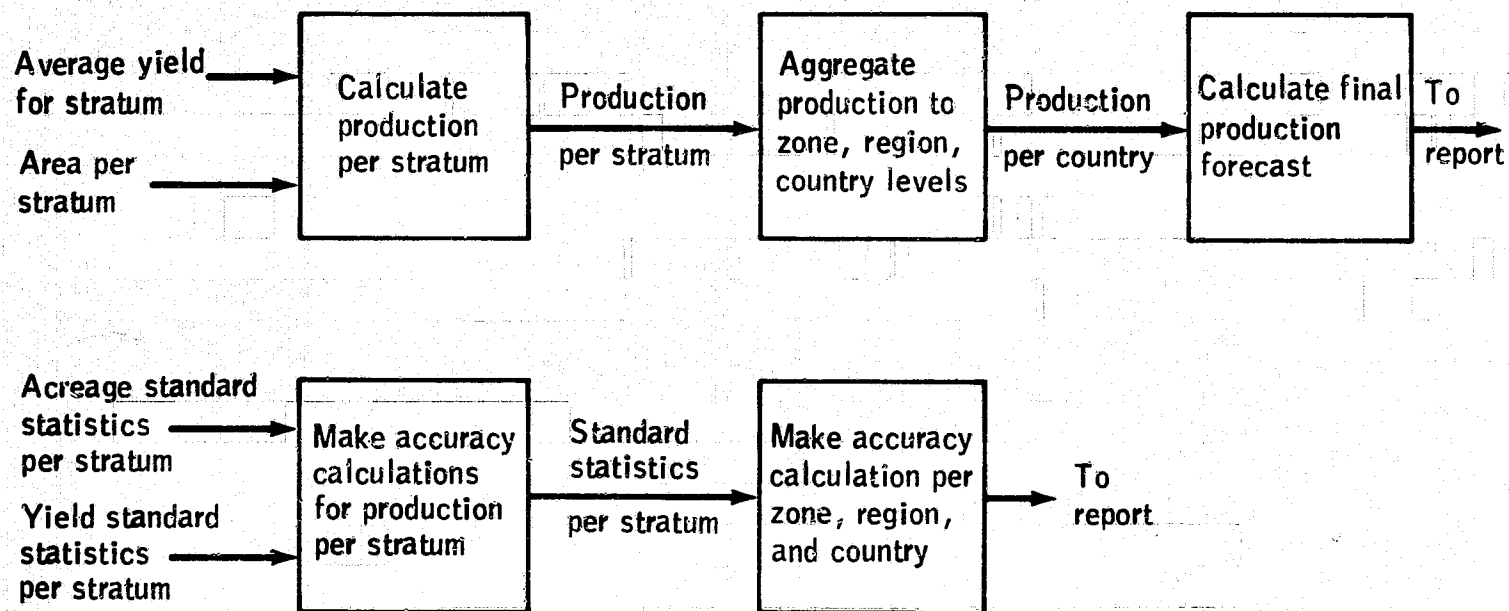


Figure 3-5.- Functional flow diagram of electronic production estimate within CAS.

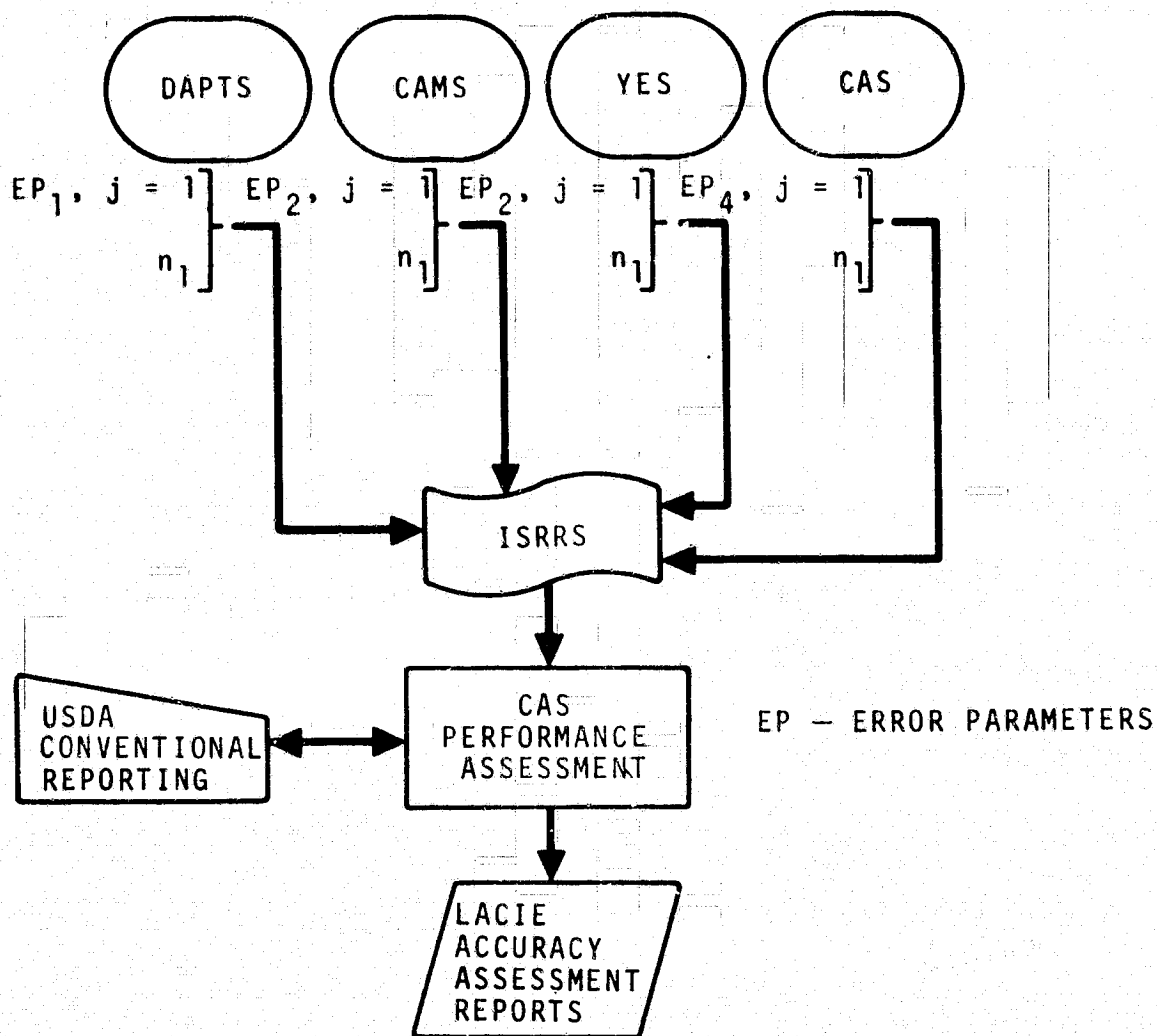


Figure 3-6.- CAS accuracy assessment full-up system operational functional flow diagram.

SECTION 4.0

REQUIREMENTS (CONSOLIDATED)

4.1 CROP ASSESSMENT

4.1.1 Full-Up System

4.1.1.1 Development.- (Required by CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB using GDSD computers). Agricultural records, Landsat mosaics, and maps for the smallest reporting subdivisions of the desired countries are used together with the effects of cloud cover as inputs to develop a sampling strategy for crop area estimation. The strategy is designed to achieve a 2 percent (or less) coefficient of variation Landsat (2 percent sample error), using analyzed data from segments 5x6 n. mi. which are placed in accordance with the constraints and throughput constraints received from Goddard and GDSD. The output products are the sample segment location, definition of each segment as spring or winter, and definitions of substrata, strata, zones, and regions. Requirements are UNIVAC 1110 and 1108. Reference Appendix D.

4.1.1.1.1 Sampling Strategy: (Required by CAS; Category 1; Implementation Responsibility: Systems Engineering/CAS-SFB). The sampling strategy developed during LACIE Phases I and II (stratification on political boundaries) as modified in Phase III will be utilized in U.S. Great Plains & USSR during Phase III & in USSR and other LACIE countries during Transition. Appendices A & B of this document detail this sampling strategy. A new sampling strategy (stratification using natural elements, referenced as the universal strata or partitioning) will be implemented in selected areas for testing during LACIE Phase III and for the Great Plains during Transition. Appendix E of this document details this new sampling strategy. These requirements are for the PDP 11/45.

~~PRECEDING PAGE BLANK~~

- 4.1.1.1.2 Aggregation procedures and software for area: (Required by CAS; Category 1; Implementation Responsibility: CAS/LACIE-EOD). Once the sampling strategy is defined and has been used to define the location of the segments, substrata, strata, zones, and regions, procedures and software will be completed to utilize the percentage of wheat from classified Landsat segments to determine crop area for each strata, zone, region, and country. Software is to be implemented on an interactive computer. (Batch processing capability is required operational on 10/1/75 and interactive capability on the PDP 11/45 on 4/1/76.)
- 4.1.1.1.3 Average yield procedures and software: (Required by CAS; Category 1; Implementation Responsibility: CAS/LACIE-EOD). Once the strata, zones, regions, and meteorological stations are defined, procedures and software can be completed to calculate average yield for each region and country. Software is to be implemented on an interactive computer.
- 4.1.1.1.4 Integration software for calculation of production from yield and area: (Required by CAS; Category 1; Implementation Responsibility: CAS/LACIE-EOD). Once the segments and strata are defined procedures and software can be completed to utilize average stratum yield and percentage of wheat from individual segments to calculate production for individual strata. Software is to be implemented on an interactive computer. (Batch processing capability is required operation on 10/1/75, and interactive capability is required operational on 4/1/76.)
- 4.1.1.1.5 Aggregation procedures and software for production: (Required by CAS; Category 1; Implementation Responsibility: CAS/LACIE-EOD). Once the sampling strategy is defined and procedures and software developed for integration of yield and area into production, procedures and software can be developed to aggregate production for each zone, region, and country. Software is to be implemented on an interactive computer. (Batch processing capability is required operational on 10/1/75 and interactive capability operation 4/1/76.)

- 4.1.1.1.6 Ancillary information: (Required by CAS; Category 3; Implementation Responsibility: CAS/USDA). Devise procedures for modifying yield, area, and production estimates interactively using information synthesis techniques to incorporate ancillary information. Information such as FAS attache reports, FAS country analyst reports, weather-related information on insects, crop disease, crop-management factors, and other TBD reports will be incorporated into the area and production forecasts.
- 4.1.1.1.7 Error estimates: (Required by CAS; Category 1; Implementation Responsibility: CAS/LACIE-EOD). Procedures and software will be designed to calculate standard statistics of area, yield, and production estimates using the ratio of true county agricultural area, estimated error in percent wheat from CAMS, standard statistics from YES, historical wheat area, yield and production for stratum, zone, region, and country, area stratum ID, bio-phase, and date of acquisition. (Batch processing capability is required operational on 10/1/75 and interactive capability on 4/1/76.)
- 4.1.1.1.8 Modify crop area, yield, and production software: (Required by CAS; Category 3; Implementation Responsibility: CAS/LACIE-EOD). Modify area yield, and production estimating software and procedures, if required, based on the results obtained in Phases I and II.
- 4.1.1.1.9 Devise procedures for choosing sample segments: (Required by CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Procedures will be defined and documented for choosing sample segments based on the sample strategy defined in 4.1.1.1. These step-by-step procedures and equipment are defined to a level of detail that will enable persons with photointerpretation, country analysts, and mensuration backgrounds to carry out all required functions. See Appendix B.

- 4.1.1.1.10 USDA Compatible Outputs: (Required by LACIE PO; Category 3; Implementation Responsibility: CAS/LACIE-ASVB). The CAS will determine the report format and content of the current crop reporting system of the USDA, and insure that the proposed reports generated by the LACIE AES are compatible. Any discrepancies will be rectified with the concurrence of the LACIE PO and USDA. The objective is to insure that the LACIE output products can be integrated into the USDA reporting system, if required, in the future.
- 4.1.1.1.11 Software to be developed: (Required by LACIE; Category 4; Implementation Responsibility: CAS/EOD). Software will be developed to use area, yield, and production for each month for years i-1, i-2, etc., to calculate area, yield, and production for year i-2, etc. for each stratum.
- 4.1.1.1.12 Test of area, yield, and production procedures and software: (Category 1; Implementation Responsibility: CAS/LACIE-FSO). The procedures and software for estimating the area, yield, and production to the country level will be tested prior to implementation in the AES. Batch processing test complete by 10/1/75 and interactive test complete by 4/1/76.
- 4.1.1.1.13 Select segments for each LACIE country: (Required by CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Sample segments are chosen using techniques described in the procedures document (4.1.1.1.10) at the rate of about 50 segments/day for a total of 4800 segments. This task is redone annually if required, and requires Landsat mosaics and maps of $1:10^6$ scale, Landsat constraints overlays, nonagricultural areas defined 5x6 n. mi. grid overlays, and 5x6 n. mi. center point overlays.
- 4.1.1.2 Operations.— The full-up system will consist of estimating wheat area, yield, and production by stratum, zone, region, and country. These parameters will be determined monthly from planting

to postharvest. Initial operations in the batch mode on the PDP 11/45 are required by 10/1/75, and full-up operation in the interactive mode on the PDP 11/45 is required 4/1/76.

- 4.1.1.2.1 Calculate total wheat area per stratum: (Required by CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). There are up to 60 sample segments (in the U.S. there is usually one segment per county) contained within each area stratum depending upon the amount of wheat grown in the area, and the stratum size. The sample segment percent wheat, correlated with the area devoted to wheat is determined from the sample segments. This is then combined with the historical area of wheat within that stratum to obtain an estimate of the current spring and winter wheat area. If drastic changes in wheat grown (as determined from the sample segments) occur, the segment will be flagged for checking, or possible reclassification by the CAMS. This operation is performed for all area strata within a country.

CAS requires the ability to aggregate CAMS classification data by six (6) subclasses (i.e., all small grains, spring small grains, winter small grains, spring wheat, winter wheat, and other), as reported in the CAMS classification summary. The spring and winter wheat subclasses are required to produce spring and winter wheat area estimates. In addition, CAMS classification results that are provided to CAS include the subclasses of small grains, winter small grains, and spring small grains. The ability to aggregate areas on a subclass basis is required and also the ability to derive the wheat area estimate when confused with small grains requires ratioing out the small grains based on the ratio of wheat to small grains from the previous year's historical data.

CAS requires the ability to aggregate data by biostage, ordering the biostages, or combinations of biostages. Biostage or growth stage information is used in crop identification on Landsat data and provides a measure for analyzing the best area identification results for CAMS. While multiple Landsat passes of equal quality exist, biostage ordering

allows for the selection of the best data for area aggregation. The CAS analyst shall have the option of aggregating by subclasses, wheat only from ratioed subclasses, and by a single biostage or group of biostages in any order up to seven (7) biostages.

Using the sample segment percent wheat (and notation in terms of suitability for aggregation), the sample segment identification, wheat growth stages, dates of acquisition, strata ID, nearest neighbors to each stratum, historical agricultural data, ancillary area data, and the area strata software, the estimates of wheat area per stratum will be calculated using interactive software and computers. Output will include historical data, report number, date of reports, strata, zone, region, country, etc. This will be done about twice per month per stratum. Initial operation required 10/1/75.

The CAMS two digit (00-99) evaluation codes (5.1.1.2.1B) determine the suitability of a segment classification for aggregation according to the following description:

- 00-09 Data is not processed by CAMS and is not used for aggregation.
- 10-19 Classification data is unsatisfactory.
- 20-29 Classification data is marginal.
- 30-39 Classification data is satisfactory.
- 40 Segment is totally nonagricultural but can be used in an aggregation.

The CAS analyst selects the numerical value of the codes or the range to be aggregated. In the case of multiple data in the same biowindow the logic should select the latest acquisition within the code threshold. In the case of codes 12, 22, and 32, which designate no significant change from last classification, the proportion classified for the segment is obtained from the last previous unsatisfactory, marginal, or satisfactory evaluation code, respectively.

- 4.1.1.2.2 Calculate standard statistics of wheat area per stratum, zone, region, and country: (Required by CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). The estimated error in percent wheat, standard deviation in the historical wheat area, estimated wheat area in each stratum, number of sample segments acquired, sample segment ID, area of strata, zone, etc; historical wheat area of strata, zone, etc; strata ID; biophase; and ratio of true county agricultural area to pseudo-county area, will be used to calculate the standard statistics of wheat area per zone, region and country. Same frequency as 4.1.1.2.1 with initial operation required 10/1/75.
- 4.1.1.2.3 Calculate area aggregation: (Required by CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Using the area aggregation software, the CAMS proportion of wheat for each strata, historical agricultural data; nearest zone neighbors; zone, region, and country descriptors; calculations are made of the wheat area for zones, region, and countries and reported and stored in the CAS and ISRRS. Each level will be done "interactively" twice per month when a significant number of the segments in an area have been analyzed. Output will also include ancillary data similar to strata report (see section 4.1.1.2.1). Each time an area report is generated, these data are then available for reporting purposes by the CAS, and no report storage is required in the computer. The data from the replaced reports are stored. These new data are treated as speculative data. Batch operation required on 10/1/75 and interactive operation on 4/1/76.
- 4.1.1.2.4 Calculate yield and production: (Required by CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). The subsystem operates a computerized crop aggregation model that will obtain a monthly estimate (beginning 4/1/75) of the average yield and production by strata, zone, region, and country. The computed yield and area are obtained from storage. Calculation of the standard statistics associated with the yield and production estimates is also required.

A. The following inputs are required:

1. Yield value of each area stratum.
2. Meteorological station coordinates - latitude and longitude in degrees and minutes. (E, W and N, S).
3. Ancillary data.
4. Yield strata boundaries.
5. Historical yield for 15 years.
6. Yield coefficients for each stratum.
7. In addition to these inputs, tracking of errors through the calculations requires the standard statistics for each yield value.

B. The yield and area estimates are integrated to obtain a production estimate for each area stratum. The yield and wheat area estimates for the strata are used to calculate production for each stratum. An average production for each stratum is calculated and modified by the ancillary data if data from that stratum are available. The production is aggregated to the zone, region, and country levels. The production at the country level is again modified based on ancillary data, if required, and then reported. Speculative data are generated.

C. The output of the CAS is designated speculative information, and access to the data and reports must be controlled. Proposed outputs are included in appendix A.

4.1.1.2.5 Calculate standard statistics of production for stratum, zone, region, and country: (Required by CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). The estimated production and yield in each stratum together with inputs required in 4.1.1.2.2 will be used to estimate the standard statistics of production at all higher levels. Same frequency as 4.1.1.2.1.

4.1.1.2.6 Software and Hardware Design Constraints: (Required by CAS, Category 1, Implementation Responsibility: CAS/LACIE-ASVB). The aggregation software will be designed to incorporate CAMS area measurements together with historical data to produce wheat area estimates for the substrata, strata, zone, region, and country. In Phase II, yield will be factored in to estimate production. The system will be operated by the CAS analyst from a CRT terminal(s) with line printer(s). The analyst will be able to operate the aggregation interactively or in batch mode. In the latter mode he will designate an area (by strata, zone, region or country), user name and user code and aggregations will be performed at all levels up to that point, with a results report being printed for that highest aggregation level. Printouts of lower level aggregations can be called for on a one-by-one or on a group basis as desired. The analyst can also generate aggregations manually, one at a time, to correct previous reports. The batch mode may then be used to aggregate upward to correct higher level reports. (Batch operation required 10/1/75 and interactive capability on 4/1/76.)

Reports listing all data in the CAS data base must be available to the CAS analysts.

All data in the DAS data base (including input) can be modified by the CAS analyst. However, each change must be accompanied by a change report which gives the time, day, month, and year, name of the person making the change, old data, new data, report changed, copy of affected report with change denoted by \$, , or * on both sides of change, and reason the change was made.

If errors are detected after a report is generated (e.g., report for Kansas for April) a capability must exist to go back and change this aggregation and generate a new report even though the error could have occurred 6 months or 1 year ago.

Normally, substrata reports and yield strata will not be called for; however, when problems occur these reports will be called for by the analyst.

4.1.1.2.7 Monitor and record: (Required by SPE; Category 2; Implementation Responsibility: CAS/LACIE-ASVB). On a weekly basis, the following information expended in the preparation of area, yield, and production estimates for each country for TBD phases.

A. Manpower: Category, number of people per category, man-hours per category, categories TBD.

B. Computer time (CPU and I/O): Designation of computer system and number of hours expended. (Implementation Responsibility: EOD develop software to track this time).

4.1.2 Interim System

Not applicable.

4.2 ACCURACY ASSESSMENT

4.2.1 Full-Up Systems

4.2.1.1 Approach.- The LACIE system will utilize remote sensor data; the location of which will be selected by a probability area sampling technique to estimate crop area, yield and production over a large area. In most previous uses of Landsat data, ground information has been used to provide known levels of accuracy. This may be impractical in LACIE where Landsat data must be processed for areas where ground information is not available. An alternate approach will be developed, which monitors various aspects of the LACIE components and through a model which interrelates these aspects, to provide an estimate of the reliability and accuracy of the LACIE outputs. This model will be verified and update continuously during the LACIE using data as acquired for the intensive study sites.

4.2.1.2 System development.- The system development is as follows.

- A. The CAS will provide an assessment of the accuracy/precision for each area, yield, and production estimate made by the LACIE using mathematical models to synthesize system behaviour. The system model will be developed by first analyzing the LACIE system and identifying possible sources of error. Initially, the LACIE requirements, procedures, and operating documents will be reviewed. In addition, monitoring of system simulation, training, and operations will also aid in identifying possible error sources. (Required by CAS; Category 3; Implementation Responsibility: CAS/LACIE-ASVB.)
- B. The importance of the various errors identified for the system will be evaluated. This evaluation will be made using data from three sources.
1. Data obtained from previous investigations into the use of Landsat data such as those obtained during the CITARS project will be evaluated. In particular, the results of the individual segment analyses and the results of the CITARS analysis of variance will be analyzed. (Required by CAS; Category 3; Implementation Responsibility: LACIE-RTEB.)
 2. A major source of data for the model development will be that obtained from the LACIE Test and Evaluation studies. A number of controlled experiments will be carried out to isolate the various components of the system, and how they individually vary relative to some associated parameters. These experiments will be designed by T&E and concurred on by the CAS. (Required by CAS; Category 3; Implementation Responsibility: LACIE-RTEB.) The results of the Test and Evaluation studies will provide the basic input for the LACIE error model development. Detailed requirements are defined in section 11.2.
 3. The magnitude of the various errors to be encountered by the system will also be

quantified by monitoring the processing of LACIE Intensive Test Site data during training, simulations, and LACIE operation. The data required to be monitored and recorded for the intensive test sites are identified in section 5.2 for each LACIE subsystem. (Required by CAS; Category 3; Implementation Responsibility: CAMS/YES/CAS/DAPTS.)

- C. The major source of error for each subsystem will be identified. The data obtained will be analyzed and a correspondence will be set up between the relevant aspects of the situation, and will be formulated into mathematical models. The models will be exercised and compared with the ground reference data for verification of the model. Also, simulated data will be processed to provide a more controlled evaluation of the mathematical models. (Required by CAS; Category 3; Implementation Responsibility: CAS/LACIEASVB for model development and GDSD for computers.)
- D. The technique and procedures required for the operational implementation of the CAS accuracy assessments will be developed, documented, and tested prior to the beginning of the LACIE full-up system. (Required by CAS; Category 3; Implementation Responsibility: CAS/LACIE-ASVB for development and GDSD for computers for testing.)

4.2.1.3

Design philosophy.- The accuracy assessment system of the CAS will be designed in a modular form to insure flexibility for modification of the subsystem without major subsystem impact. The operational mode of the accuracy assessment system is required to be interactive. (Required by CAS; Category 1; Implementation Responsibility: ISSRS.) That is, the user shall be able to address information for change and storage or retrieval, responding only to user input directive. Information security shall be maintained to prohibit unauthorized user access. (Required by CAS; Category 2;

Implementation Responsibility: ISSRS.) The design should allow for data retrieval of TBD accuracy parameters on magnetic tape, compatible with the PDP 11/45 for offline processing and analysis. (Required by CAS; Category 3; Implementation Responsibility: ISSRS.)

4.2.1.4

CAS accuracy assessment operation.- The CAS accuracy assessment operation is as follows.

- A. The CAS shall insure that each subsystem monitors and records the parameters necessary to provide an assessment of the reliability and accuracy of the area, yield, and production estimates. The parameters required for the assessment are listed in section 5.2. Additional parameters are TBD. (Required by CAS; Category 3; Implementation Responsibility: CAS/LACIE-ASVB.) These data will be processed and analyzed via the LACIE error model, and an estimate of the accuracy and reliability will be provided with all LACIE outputs. (Required by LACIE PO; Category 3; Implementation Responsibility: CAS/LACIE-ASVB for error model and EOD for computer support.)
- B. Any discrepancies in the LACIE area, yield, and production reports, identified by the USDA evaluation of the information contained in these reports, will be investigated by the CAS. If the source of the discrepancy is determined to be within the AES, the CAS will analyze the causes and report to the AES and Quality Assurance Managers, and make recommendations to the LACIE PO. (Required by LACIE PO; Category 3; Implementation Responsibility: CAS/LACIE-ASVB.)

4.2.2

Interim System

Not applicable.

SECTION 5.0

INPUT REQUIREMENTS

Inputs required by CAS from the other LACIE functional elements are defined in the following subparagraphs.

5.1 CROP ASSESSMENT

5.1.1 Full-Up System

5.1.1.1 DAPTS.- The following requirements are applicable to DAPTS.

A. Landsat data requirements

1. Landsat mosaics (Category 1; Implementation Responsibility: DAPTS). Uncontrolled Landsat-2 mosaics of MSS-5 for "green phase" of wheat growth stage a 1:1,000,000 scale of the Phase II LACIE regions are required 60 days prior to the start of acquisition of Landsat data for Phase II. Full frame Landsat-2 color IR transparencies of the eight LACIE countries.

B. Field data requirements

Not applicable.

C. Historical agricultural data requirements

1. Historical Agricultural Statistics (Category 1; Implementation Responsibility: DAPTS). Total area, total agricultural area, wheat area (planted and harvested), yield, and production data for the smallest reporting political subdivision for each LACIE country for the last 15 years (if available) are required. These data are required on computer compatible magnetic tape, one tape per country per year, updated on an annual

basis. Delivery date for U.S. is ASAP. Delivery of other LACIE countries' data is required by 11/1/75. Specific data requirements and format are described in NASA/USDA ICD, Vol II.

2. Overlap regions (Category 1; Implementation Responsibility: DAPTS). Map overlays for LACIE site location maps of 1:1,000,000 showing the winter/spring wheat overlap regions for the U.S., USSR, and Peoples Republic of China (PRC) are required. These overlays will be reviewed annually and updated as required. The criterion for an update is whenever overlap boundaries change. These overlays are required ASAP for the U.S. The overlays for the USSR and PRC are required by August 15, 1975. One copy of each overlay is required.
3. Boundaries of political subdivisions (Category 1; Implementation Responsibility: DAPTS). Four copies of maps of 1:5,000,000 scale or larger of each LACIE country delineating the boundaries of the smallest political subdivisions, are required. The maps must be of known projection. The maps shall cover each country in detail at least as fine as the data presented in 5.1.1.1 C.1. The smallest political subdivisions for acreage aggregation are as follows:

<u>Country</u>	<u>Political Subdivision</u>
Argentina	Partido
Australia	Divisions
Brazil	Municipality
Canada	Mixture of divisions and subdivisions
India	District
PRC	Provinces
U.S.	Crop Reporting District
USSR	Oblast

These maps are required as specified in paragraph 5.1.1.1 C.1. The maps will be

updated annually, when more detailed data become available, or when boundary changes occur.

4. Cropping practices (Category 1; Implementation Responsibility: DAPTS). Cropping practices reports (ancillary summaries) are required for each stratum identified in paragraph 5.1.1.1 C.3. The reports produced for YES are acceptable.
5. Crop calendars (Category 1; Implementation Responsibility: DAPTS). One copy of mean historical crop calendars based on last 15 years, if available, for each acreage stratum identified in paragraph 5.1.1.1 C.3 is required. These crop calendars will be used in determining the acquisition windows for Landsat. The calendars will be updated annually.
6. Agricultural reports (Category 1; Implementation Responsibility: DAPTS). USDA reports containing current information on wheat area, yield, and production in all LACIE countries are required. Reports published by LACIE foreign countries are also acceptable. One copy of the following reports is required:
 - a. Annual summary of area, yield, and production of wheat for each LACIE country.
 - b. Foreign agriculture circular - grains (FAS) quarterly.
 - c. Foreign agriculture (FAS) - weekly.
 - d. Wheat situation (ERS/SRS) - Quarterly.
 - e. Foreign agricultural attache reports - monthly.
 - f. World agricultural and trade (FAS) - quarterly.

- g. U.S. State publication in cooperation with SRS giving area, yield, and production at the country level, yearly. Required for all states containing sample segments.

5.1.1.2 CAMS.-- The following information applies to the classification and mensuration subsystem.

5.1.1.2.1 Electronic:

- A. Ready-access electronic outputs for sample segments (Category 1; Implementation Responsibility: CAMS). The following parameters shall be stored on magnetic tape for each classification run. For the recognition segment:
 - 1. Sample segment number
 - 2. DPR number
 - 3. Run 10
 - 4. Bhattacharyya separability
 - a. All channels
 - b. Channel subset used
 - 5. External statistics tape number
 - 6. Number of recognition segment subclasses
 - 7. Number of merged image fields
 - 8. Number of pixels in D/O fields
 - 9. Number of pixels in D/U fields
 - 10. Date of classification
 - 11. Number of clustering runs
 - 12. Number of chaining runs
 - 13. Number of acquisitions used for classification

14. Channels used in classification
15. Current biological windows
16. Segment type identifier (spring/winter)
17. Sample segment use indicator (training or not training)
18. Dates of acquisition on merged image
19. Segment number, segment type indicator, and acquisition dates for each training segment
20. Threshold and a priori values used in classification (by category)
21. Subclass classification results
 - a. Subclass name
 - b. Statistics manipulation flags
 - c. Pixels classified into subclass
 - d. Pixels threshold from subclass

For each field in segment:

22. Field name
23. Subclass name
24. Field type
25. Field population
26. Number of subclasses
27. Name of each subclass
28. Mean level adjustment, Sun angle correction flags for that (sub)class

29. Pixels classified into that subclass

30. Pixels threshold from that subclass

This tape shall be stored by ISRRS in the physical library for use by CAMS and CAS.

B. The analyst evaluation of the recognition segment shall be manually passed to CAS and stored in the physical library by ISRRS. The following evaluation codes shall be used:

Code 01: Not machine processed - clouds, haze, snow, etc. Means that the segment cannot be processed through the system because clouds, haze, etc., make interpretation and analysis impossible.

Code 02: Not machine processed - confusion crops or other interpretation problems.

This code should be used when a segment cannot be processed because of interpretation difficulties, especially when confusion between wheat/small grains and other crops is such that a wheat/small grains estimate cannot be determined. The rules for such a decision are to be negotiated between CAMS and CAS and included in the CAMS Detailed Procedures.

Code 03: Bad data - due to technical problems, not reordered.

In this case the segment cannot be processed due to technical problems arising from unsatisfactory histogram, excessive scanline drop, etc. If the segment is not reordered, this code is used.

NOTE: If the segment is re-ordered, no code should be listed and no data sheet should be passed to CAS since this segment will again be sent through the CAMS system for evaluation. Any segment that cannot be processed due to clouds, etc., or technical problems, should be passed as 01 or 03.

Code 05: Not machine processed - dormancy.

In cases where recognition is a problem due to the crop being in a state of dormancy this code is used.

Code 07: Not machine processed - preemergence.

In cases where the acquisition is prior to the criteria established for fall wheat recognition (TBD) this code is used.

Code 09: Not machine processed - multiple acquisitions.

Code 09 was developed to take care of the problem of processing multiple acquisitions of a segment at the same time. When multiple acquisitions are available at one time, only one segment, which is determined to be the better acquisition for interpretation and analysis by the CAMS analyst, is selected for complete processing. Other acquisitions are listed as Code 09 to indicate that they had been reviewed but were not processed through the system. This code enables CAS to review the segment results and account for

each acquisition and how it was evaluated by CAMS.

Code 10: Unsatisfactory - unsatisfactory results for segment.

This code is to be used for any acquisition that processed through the system and, based on CAMS evaluation procedures, designated unsatisfactory.

Code 12: Unsatisfactory - no significant change.

This code is used when the new acquisition is evaluated to have no significant change from the previous unsatisfactory evaluation for the segment.

Code 14: Unsatisfactory - rework, reevaluated segment.

This code is used when a segment which was previously passed to CAS is reworked.

Code 18: Unsatisfactory - machine processed multitemporal analysis.

This code is used when more than one acquisition date is used to produce an unsatisfactory proportion estimate.

NOTE: All acquisition dates used in processing should be listed on the CAMS Evaluation Form.

Code 20: Marginal - marginal results for segment.

This code is to be used for any acquisition that processed through the system and, based on CAMS evaluation procedures, designated marginal.

Code 22: Marginal - no significant change.

This code is to be used when the new acquisition is evaluated to have no significant change from the previous marginal evaluation for the segment.

Code 24: Marginal - rework, reevaluated segment.

This code is used when a segment which was previously passed to CAS is reworked.

Code 28: Marginal - machine processed multitemporal analysis.

This code is used when more than one acquisition date is used to produce a marginal proportion estimate.

NOTE: All acquisition dates used in processing should be listed on the CAMS Evaluation Form.

Code 30: Satisfactory - satisfactory results for segment.

This code is to be used for any acquisition that processed through the system and, based on CAMS evaluation procedures, designated satisfactory.

Code 32: Satisfactory - no significant change.

This code is to be used when the new acquisition is evaluated to have no significant change from the previous satisfactory evaluation for the segment.

Code 34: Satisfactory - rework, reevaluated segment.

This code is used when a segment which was previously passed to CAS is reworked.

Code 36: Satisfactory - less than 5% manually (hand) counted.

This code is to be used for any segment in which the proportion estimate is manually counted rather than machine processed.

NOTE: This category would not be used if the segment was a rework segment. Code 34 should be used.

Segments in an agricultural area that have a 0% proportion estimate are to be designated Code 36.

Code 38: Satisfactory - machine processed multitemporal analysis.

This code is used when more than one acquisition date is used to produce a satisfactory proportion estimate.

Code 40: Segment is totally nonagricultural.

This code is used when the segment is evaluated as having no agriculture at all, i.e., no discernible field patterns.

NOTE: Segments in an agricultural area that have a 0% proportion estimate are to be designated Code 36.

5.1.1.3 YES.- The following information pertains to the yield estimation subsystem.

5.1.1.3.1 Electronic:

A. Meteorological data. (Required by CAS: Category 1; Implementation Responsibility: YES/NOAA). Current standard meteorological data is required for use by the CAS analysis in accounting for episodal weather events in crop aggregations and for analysis of crop production.

These records should include, as a minimum, monthly average of daily maximum and minimum temperatures and precipitation amounts for the periods of wheat yield records in that area. Where available, the record may contain more temporal detail and other parameters. Examples of parameters that may be available are average daily maximum temperature, average daily minimum temperature, monthly mean rainfall, daily rainfall reports, snowcover, drought index, and solar radiation.

Meteorological stations from which data are required are all those as available in each smallest political subdivision containing one or more LACIE sample segments. It is anticipated that up to 5000 meteorological stations may be involved. These records are required in either published volumes or on computer-compatible magnetic tape.

Location of meteorological stations latitude and longitude in degrees and minutes. Yield strata descriptors. Yield meteorological stations correlated with strata.

Once monthly NOAA will deliver to NASA-JSC-NOAA/YES the daily transmission plus a monthly summary of the WMO data received and transmitted daily. This is required on magnetic tape in the formats of Appendixes F and G of Vol. III.

- B. Yield coefficients (Category 1; Implementation Responsibility: YES). Provide yield coefficients for each strata and each truncation for spring and winter wheat for all LACIE countries using the current CCEA-1 yield model prior to model operations. Provide MAP factors using Feyerherm (KSU) yield model for natural strata in Kansas and North Dakota in the U.S. Great Plains and for natural strata in Tselinograd and Kurgan oblasts in the USSR for use during Phase III, for natural strata in all of the LACIE U.S. Great Plains area for use during Transition.
- C. Yield values (Category 1; Implementation Responsibility: YES). Provide monthly spring and winter wheat yield values and standard statistics for each stratum (natural or political as appropriate) using both the CCEA-1 and the Feyerherm yield models for areas specified above. The standard statistics will show the relative contribution of historic and current meteorological data. These data are to be provided as defined in NOAA/NASA/USDA ICD (LACIE C00710).
- D. Terminal (Category 1; Implementation Responsibility: YES for the Page Bldg. interface to the phone lines, FSO for phone lines and terminal at JSC). A read-only terminal for pursuing the YES data base and predicted yield values shall be provided at JSC connected to the YES prediction computer in Washington, D.C. and Columbia, Md. Data rates shall be low enough to permit use of acoustic coupler. Use shall be 2 hours per day, 7 days per week. YES shall provide personnel to operate JSC terminal. YES shall determine where the terminal shall be located.

5.1.1.3.2 Nonelectronic:

- A. Crop calendars (Category 1; Implementation Responsibility: YES). Crop calendars and real-time adjustments to crop calendars by strata provided in season approximately once per two weeks as defined in the NOAA-NASA-USDA ICD.

- B. Meteorological summary (Category 1; Implementation Responsibility: YES). Tabulations, overlay maps, and narratives will be provided for an overall view of the prevailing meteorological conditions and annotations, whether normal, above normal, or below normal, and to what extent. Flags will be provided to CAS for aggregate consideration on the amount of stratum, zone, region, or country which has experienced abnormal weather (unseasonably high or low temperatures, precipitation, storms, etc.). Delivery requirements are defined in the NOAA-NASA-USDA ICD.
- C. Yield strata boundaries (Category 1; Implementation Responsibility: YES). Boundaries of yield strata in all LACIE countries drawn on from 1:250,000 to 1:10⁶ ONC maps as appropriate. Annual with monthly corrections, if required.

5.1.1.4 CAS.- The following information pertains to the crop assessment subsystem.

5.1.1.4.1 Sample strategy:

- A. Selection of sample strategy for use in Phase I (Category 1; Implementation Responsibility: CAS/LACIE-ASVB). See Appendix C.
- B. Select sample segments (Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Allocation of sample segments with definition of center coordinates (latitude and longitude in degrees and minutes) and number/stratum and ID for each acreage stratum. Each segment will be identified as spring or winter wheat. Sample segment ID and allocation for LACIE countries.

<u>Country</u>	<u>Segments</u>	<u>Allocated ID numbers</u>
Brazil	47	001-099
Argentina	165	500-799
U.S. (New Sam- pling Strategy)	164	800-999
United States	601	1000-1999
India	626	2000-2999
Canada	283	3000-3499
Australia	257	4000-4499
China	810	5000-6499
USSR (New Sam- pling Strategy)	80	6500-6999
USSR	1947	7000-9499

- C. Pseudo substrata (Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Number of 5x6 nm rectangles per substratum minus excluded nonagricultural areas.
- D. Substratum ratios (Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Ratio of true substratum area to pseudo substratum area.
- E. Test sample strategy (Category 1; Implementation Responsibility: CAS/LACIE-ASVB). The sample strategy will be tested prior to implementation in the AES.
- F. Subdivision definition (Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Definition of up to 4800 substrata, 400 strata, 150 zones and 50 regions with boundaries and ID.
- G. Neighbor strata (Category 1; Implementation Responsibility: CAS/LACIE-ASVB). List of nearest neighbor strata for each area stratum and substratum.
- H. Historical wheat acreage (Category 2; Implementation Responsibility: CAS/LACIE-ASVB). Historical wheat area and total area for each area substratum in each LACIE country for inclusion in the monthly report for comparison.

- I. Historical cloud cover (Category 1; Implementation Responsibility: CAS/ASVB). Historical cloud cover statistics of wheat growing areas (>2000 acres wheat) in LACIE countries. Monthly unconditional probabilities are needed for each acreage stratum of the LACIE countries. An overall assessment of the probability of acquiring the stage 1, 2, 3, 4; 1, 2, 3; 1, 3, 4; 2, 3, 4; 2, 3, 4; etc. for each stratum is required.
- J. Index of sample segments (Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Index substrata to 400 acreage strata, index of sample segments to 4800 substrata, index of area strata to 150 zones, index of zones to 50 regions, and index of regions to 8 countries.

5.1.1.5

ISRRS. - The following information applies to the information storage, retrieval, and reformatting subsystem.

- A. The ISRRS shall supply to CAS sample segment classification reformatted from the results tape generated by CAMS at 1-day intervals. ISRRS shall also merge the daily tapes into one master tape for delivery to CAS. The nonelectronic requirements of section 5.0 and 6.0 require library-type storage by the ISRRS. (Category 1; Implementation Responsibility: ISRRS).
- B. Sample segment status matrix (Category 1; Implementation Responsibility: ISRRS.) Sample segment status report (in tabular format) required daily. Status of 4800 sample segments shown as date Landsat data acquired, wheat biological stage, acquired - not classified - acquired - classified suitable, acquired - classified marginal, acquired - classified unsuitable, acquired - backlogged, not acquired. Computer capability to sort segment status by country, region, and zone and by classification - suitable, marginal and unsuitable is required in Phase II and Phase III. Sort capability is Category 2.

5.1.1.6

SPE.-

5.1.1.7

IE.- (Category 3; Implementation Responsibility: IE/USDA). The type, content, format, and frequency of information required from the LACIE AES.

5.1.1.8

Research.- (Category 1; Implementation Responsibility: LACIE-RTEB). All reports concerning the research specified in section 10.0.

5.1.1.9

Test and evaluation.- (Category 1; Implementation Responsibility: LACIE-RTEB). LACIE test and evaluation reports as they are available. These reports contain the results of the efforts in the LACIE test and evaluation of intensive test sites, ASCS test sites, etc.

5.1.2

Interim Systems

Not applicable.

5.2

ACCURACY ASSESSMENT

Inputs required by the CAS for the system accuracy assessment from the other LACIE functional elements are defined in the following subparagraphs.

5.2.1

Full-Up System

5.2.1.1

DAPTS.- The following information applies to the DAPTS.

A. Landsat data requirements

1. Landsat MSS calibration documentation (Category 1: Implementation Responsibility: DAPTS). Documentation is required on the GSFC calibration procedures for LACIE data; in particular, the procedure for replacement of data due to

out-of-sync conditions and the relationship of the radiometric error versus the signal strength of spectral energy level.

2. Landsat MSS registration documentation (Category 1: Implementation Responsibility: DAPTS). Documentation is required on the GSFC registration procedures for LACIE; in particular, the definitive procedures used in both the automatic and manual modes, the error analysis study supporting the registration accuracy requirements and a further definition of the origin of the header record information pertaining to registration.
3. Registration accuracy report (Category 1; Implementation Responsibility: DAPTS). The DAPTS is required to determine the GSFC registration accuracy on a periodic basis for a number of the LACIE segments needed to make a valid statistical inference of the registration accuracy. A semiannual report is required which presents the results of the registration accuracy analysis.

B. Field data requirements

1. Ground truth (Required by CAS; Category 1; Implementation Responsibility: DAPTS/S&AD-FSO/USDA).

DAPTS will provide the following data for each LACIE Intensive Test Site.

- A. One copy of a 1:24,000 scale rectified B&W photomosaic of each ITS. The photomosaic is to be made from high altitude aerial photography acquired once per year, just before wheat harvest.
- B. One copy of a 1:24,000 scale transparent overlay registered to the photomosaic of (a) above and delineating the boundaries of each field in the test site. Each field will be numbered and a tabular listing provided containing the following data:

1. Field number.
2. Field acreage.
3. Crop and variety being grown.
4. Is field irrigated? If so, how?
5. Is field fertilized? If so, how?
6. Planting date.

These data are to be acquired once per crop year, after spring planting and before wheat harvest.

- C. One copy of a 1:24,000 scale field boundary map registered to the photomosaic of A above and delineating the boundaries of only those fields containing fall-planted crops in each Intensive Test Site in Texas, Kansas, Indiana, and Washington. Each field will be numbered and a tabular listing provided containing the information of item B above. These data are to be acquired once per crop year, after fall planting and before the beginning of dormancy.
- D. One copy of periodic observation data acquired at 18-day intervals from approximately 50 fields in each Intensive Test Site. Fields selected for observations must contain both wheat and other crops grown in the area. Fields selected must be 40 acres or larger in size. Adjacent fields planted in the same crop and divided by fences or farm roads may be grouped to obtain a 40 acre or larger area, if necessary. Data are required from the end of dormancy for fall planted crops or spring crop plantings until harvest. Data are to be acquired concurrent with Landsat-2 overpasses. The following data are required in tabular form:
 1. Growth stage
 2. Percent ground cover
 3. Plant height
 4. Surface moisture

5. Weed growth
 6. Field operations
 7. Growth/yield detractants
 8. Stand quality rating
 9. Vertical color photos
 10. Observation date
 11. General comments on field and test site conditions
- C. Historical agricultural data requirements.
Same as 5.1.1.1 C.
 - D. Real-time meteorological data requirements.
Not applicable.
 - E. Historical meteorological data requirements.
Not applicable.

5.2.1.2

CAMS - classification summary report.- The following items are required from the CAMS summary report:

- A. AI misidentification errors (Category 3; Implementation Responsibility: CAMS). An estimate of the misidentification errors due to AI, per training segment and stratum.
- B. Cluster procedure identification (Category 1; Implementation Responsibility: CAMS). Documentation of whether clustering was used or not for each segment processed. Total classes available and the number of subclasses obtained from clustering and/or used in the classification for each class will be recorded.
- C. Signature extension parameters (Category 1; Implementation Responsibility: CAMS). Specify the distance, direction, time-phase (if different), and the Sun angle and mean level corrections (if applied) for the segment to

which signature extension was carried out. Also, specify the segment ID numbers to which, and from which, the signature extension was done.

D. Classification parameters

1. The number of channels employed in each classification; specifically the bands for each temporal period selected. (Category 1; Implementation Responsibility: CAMS).
2. The a priori probabilities used in each classification, if unequal. (Category 1; Implementation Responsibility: CAMS).
3. Classification accuracies for the training fields data sets for all segments for which AI data is obtained. (Category 2; Implementation Responsibility: CAMS).
4. Interclass and intraclass Bhattacharrya distance associated with the features for each segment processed. (Category 1; Implementation Responsibility: CAMS).
5. Classification bias for each sample segment. (Commission-omission error). (Category 1; Implementation Responsibility: CAMS).
6. Classification results for each intensive study site and the AI fields for the site. Also, the identification (in terms of coordinates) of each AI field. (Category 1; Implementation Responsibility: CAMS).

5.2.1.3 YES.- The following information is applicable to YES.

- A. Yield accuracy (Category 1; Implementation Responsibility: YES). The YES will provide yield estimates for each yield strata, associated yield contribution of each coefficient, prediction error for each strata estimate of yield, and the WMO number of each weather station used for each yield stratum and the

relative weight given to the weather from each weather station in making the yield estimate for that stratum. These data will be provided monthly for LACIE areas specified using yield models specified for those areas.

- B. Interstrata correlation (Category 1; Implementation Responsibility: YES). Provides an estimate of the interstrata yield correlation statistics.

5.2.1.4 CAS.-- The following information pertains to CAS.

- A. Aggregation parameters (Category 1; Implementation Responsibility: CAS/LACIE-ASVB). The CAS will provide for each stratum, region, zone, or country, aggregated, the following information:
 - 1. Number of sample segments expected and used per stratum/substratum.
 - 2. Number of passes and which biological phases were used for each sample segment.
 - 3. Historical data wheat area estimates for each stratum, region, zone, and country, and the accuracy of these estimates.
 - 4. Crop inventory estimates (area yield and production) in percentage of pixels by sample segment, stratum, region, zone and/or country.
 - 5. Estimate of the variance in the wheat area estimates for each stratum, region, zone, etc.
 - 6. The relationship obtained by comparing an estimate with a corresponding historical estimate.
- B. Classification and AI training-field accuracies compared on a pixel-by-pixel basis for all intensive study site areas within a 5x6 mile segment as compared to ground truth

data tapes. (Category 1; Implementation Responsibility: CAS/LACIE-ASVB).

5.2.1.5 ISRRS.- (Category 2; Implementation Responsibility: ISRRS). The ISRRS will provide storage and retrieval and provide to CAS on a demand basis the input physical data requirements identified in section 5.2.

5.2.1.6 SPE. - None identified.

5.2.1.7 IE - LACIE Information Evaluation Reports.- (Category 3; Implementation Responsibilities: IE). The information evaluation will provide the CAS with LACIE output evaluation 30 days after receipt of the LACIE reports. In addition, any IF accuracy assessment requirements shall be identified to the CAS.

5.2.1.8 Research.- Research efforts are required for a quantitative assessment of each technique to be developed and implemented in the LACIE system. Listed in section 10.2 are areas in which investigations are required for the assessment of LACIE full-up system.

5.2.1.9 Test and evaluation.- The following information pertains to test and evaluation.

- A. Testing documentation (Category 1; Implementation Responsibility: LACIE-RT&EB). The results of all subsystem testing will be documented and provided to CAS. Any interim results should be provided as available.
- B. Error Model Evaluation (Category 1; Implementation Responsibility: LACIE-RT&EB). Test and evaluation of all error models developed and/or utilized by the CAS.

- C. CITARS results (Category 1; Implementation Responsibility: LACIE-RT&EB). The results of the CITARS segment analysis, and the results of analysis of variance, shall be supplied to CAS-AA.

5.2.2

Interim System

Not applicable.

SECTION 6.0

OUTPUT PRODUCTS

Outputs from CAS which will be provided to the other LACIE functional elements are defined in the following subparagraphs.

6.1 CROP ASSESSMENT

6.1.1 Full-Up System

6.1.1.1 DAPTS.- The following information applies to DAPTS.

- A. List of sample segment numbers and coordinates (Computer Card Coding Sheets). (Required by: DAPTS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Implement the LACIE sampling strategy and select and provide to DAPTS the geographic center coordinates (latitude and longitude) of 4800 numbered sample segments. The coordinates will be specified in degrees and minutes (to the nearest minute) of latitude and longitude. Selected sites will have been judged against all system constraints (see section 4.1.1) and found acceptable. The data will be supplied to DAPTS through entry on computer card coding forms in the format specified in the JSC card format documentation.
- B. GSFC data order (computer card coding sheets). (Required by: DAPTS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). The CAS will prepare computer coding sheets containing all sample segment data needed to activate the GSFC acquisition and preprocessing of ERTS data. These coding sheets will be prepared in the format specified by the JSC card format documentation and will contain such information as the segment number, location, training site designation, definition of biological windows, and color code information

for the PFC. Inputs will be obtained from CAMS, YES, and other subsystems as appropriate. Completed forms will be provided to DAPTS.

6.1.1.2

CAMS.-- The following information is required by CAMS.

- A. Sample segment location (Required by: CAMS; Category 1; Implementation Responsibility: CAMS/LACIE-ASVB). Sample segment center coordinates - latitude and longitude in degrees and minutes. The nominal location of all sample segments shall be indicated on transparent overlays (three copies) suitable for registration upon USAF ONC's (Operational Navigational Charts) or Landsat uncontrolled mosaics. These overlays shall indicate the achieved Landsat-2 ground tracks, LACIE frame numbers, and LACIE overlap zones.
- B. Sample segment histograms (Required by: CAMS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Histograms indicating the distribution of number of sample segments per unit area for spring and winter wheat regions of each country.
- C. Landsat acquisition rates (Required by CAMS; category 1; Implementation Responsibility: CAS/LACIE-ASVB). Sample segment Landsat data acquisition rates for spring and winter wheat in segments per month by country. The individual segment numbers for each predicted acquisition date during the wheat phenological development stage.
- D. Flagged area output (Required by CAMS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Sample segment classification that results in more than 20 percent change in the previous biophase percentage wheat area will be returned to CAMS for classification verification.

6.1.1.3

YES.- The following information is required by YES.

- A. Yield prediction and Production estimation feedback (Required by: YES; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). The CAS will provide to YES flagged outputs concerning yield and/or production estimation that varies +15 percent from the historical mean, or departing significantly from surrounding strata.
- B. Definition of strata for yield prediction (Required by: YES; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). The CAS will provide a definition of the strata for which yield forecasts must be supplied to CAS.
- C. Accuracy assessment (Required by: YES; Category 1; Implementation Responsibility: CAS LACIE-ASVB). Provide accuracy assessment plans and procedures that specify those tasks for which YES is responsible.

6.1.1.4

CAS.- The following information is required by CAS.

- A. Area, yield, and production reports (Required by: CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Area, yield, and production estimates with standard statistics for spring and winter wheat for each stratum, zone, region, and country in the LACIE. The date, and report numbers are on each estimate. Stratum production will be stratum acreage times stratum yield. The units of the standard reports for wheat area, yield, and production will be hectares quintals/hectare, and thousand metric tons, respectively. Additional reports for the U.S. and Canada will be prepared in acres, bushels/acre, and thousand bushels. The area reports at each level will contain the number of each biostage used to prepare the report. The lowest level report will contain the date of the acquisition of each segment(s) used to prepare the report

and whether it is training or recognition, and multitemporal or not. Room for 7 biostages of classification results must be provided for and each report must list the years of historical data in the data base. See Appendix A for details of the reports. These reports are considered speculative information.

- B. Sample segment report (Required by: CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Number and ID of aggregatable sample segments determined to be satisfactory, marginal, or unsatisfactory (by CAMS using the evaluation codes described in Section 5.1.1.2.1B) and percentage of sample segment pixels classified as wheat, other, unidentifiable, and threshold; estimated error, biophases, and acquisition date for each area substratum allocated for area aggregation are reported weekly.
- C. Satisfactory segment area estimate for each substratum (Required by: CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Spring and winter wheat area estimate, standard deviation, and percentage of mean based on 15 years for each stratum using satisfactory segments. Satisfactory segments are segments whose evaluation codes (5.1.1.2.1B) are between 30-39.
- D. Satisfactory plus marginal area estimate for each substratum (Required by: CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Spring and winter wheat area estimate, standard deviation, and percentage of mean based on 15 years for each stratum using satisfactory plus marginal segments. Marginal segments are those segments whose evaluation codes (5.1.1.2.1B) are between 20-29.
- E. Nearest neighbors area estimate for each substratum (Required by: CAS; Category 2; Implementation Responsibility: CAS/LACIE-ASVB). Spring and winter wheat area estimate, standard deviation, and percentage of mean based on 15 years for each substratum using nearest neighbors substratum.

- F. Manual selection area estimate (Required by: CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Spring and winter wheat area estimate, standard deviation, and percentage of mean based on 15 years for each stratum using manual selection of suitable segments.
- G. Manually deleted sample segments (Required by: CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). ID's of sample segments manually deleted from area estimation and ID's of neighboring strata used for area estimation.
- H. PPS area strata (Required by: CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Numbers and ID's of selected and non-selected PPS area strata for spring and winter wheat in 150 zones.
- I. PPS strata area estimates (Required by: CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Spring and winter wheat estimated area, standard deviation, and percentage of mean based on 15 years for each strata based on PPS substrata.
- J. Region and country area estimates (Required by: CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Spring and winter wheat estimated area, standard deviation, and percentage of mean based on 15 years for each area region and country.
- K. Yield estimates (Required by: CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Spring and winter wheat yield, yield variance, and percentage of mean yield based on 15 years for each area stratum, zone, region, and country in LACIE, with yield report date and number. Speculative data.
- L. Spring and winter wheat production (Required by: CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Spring and winter wheat production, production variance, and percentage of mean production based on 15 years for each area stratum, zone, region, and country in LACIE. Speculative data.

- M. Total wheat production (Required by: CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Total of spring and winter wheat production, standard deviation, percentage of mean production based on 15 years for each area stratum, zone, region, and country in LACIE. Speculative data.
- N. Sampling strategy documentation (Required by: CAS; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). Documentation on the sampling strategies used for each country and the estimated sampling error for each strategy as a function of the input data.

6.1.1.5 ISRRS.- Not applicable.

6.1.1.6 SPE.- (Required by: SPE; Category 2; Implementation Responsibility: CAS/LACIE-ASVB responsible for reporting and GDSD for software to track computer usage times). Monitor and record TBD throughput and resource information of CAS operations for a TBD period of time on an as-requested basis.

6.1.1.7 IE.- (Required by: IE; Category 1; Implementation Responsibility: CAS/LACIE-ASVB). The CAS will submit CAS Monthly Crop Reports (CMR), CAS Un-scheduled Reports (CUR), and CAS Annual Reports (CAR) to IE, Washington, DC. The contents, format and schedule of these reports will be in accordance with Section 4.0 of the Crop Assessment Subsystem/Information Evaluation (CAS/IE) Interface Control Document (LACIE-C00709).

6.1.1.8 Research.- (Required by Research; Category 2; Implementation Responsibility: CAS/LACIE-ASVB). Computation of cloud cover over intensive test sites.

6.1.1.9 Test and evaluation.- Not applicable.

6.1.2 Interim System

Not applicable.

6.2 ACCURACY ASSESSMENT

Outputs from CAS accuracy assessment to be provided to the LACIE functional elements are defined in the following subparagraphs.

6.2.1 Full-Up System

6.2.1.1 Accuracy estimates.- (Required by: IE; Category 2; Implementation Responsibility: CAS/LACIE-ASVB). The accuracy assessment outputs will be included as part of the standard CAS outputs as required by the IE. The information to be included will be an estimate of the accuracy and reliability of the LACIE output.

6.2.1.2 Discrepancy evaluation.- (Required by: IE; Category 2; Implementation Responsibility: CAS/LACIE-ASVB). The CAS will investigate any discrepancies documented in the periodic USDA evaluation. If the source of the discrepancy is determined to be within the AES, the CAS will analyze and report the causes and make recommendations to the AES and quality assurance managers.

6.2.1.3 System testing.- (Required by: CAS; Category 4; Implementation Responsibility: DAPTS/CAMS/YES/CAS and ISRRS). A set of TBD test data will be supplied to DAPTS, CAMS, YES, CAS, and ISRRS.

6.2.2 Interim Systems

Not applicable.

SECTION 7.0

INTERFACE REQUIREMENTS

The interface requirements of CAS will be identified as the LACIE subsystem requirements documents are baselined. Interface control documents will be generated 90 days after baselining the requirements. The current ICD requirements identified are CAS-IE/USDA - output reports and feedback from USDA.

PRECEDING PAGE BLANK NOT FILLED

SECTION 8.0

OPERATIONAL REQUIREMENTS AFFECTING THE DESIGN

8.1 CROP ASSESSMENT

8.1.1 Throughput Requirements (Req'd by: CAS; Cat. 1;
Impl. Resp.: CAS/LACIE-ASVB)

There are 4800 sample segments classified four times per wheat growing season. Sample segment throughput requirements per country are TBD. Calculate the area, average yield, and production for 4800 substrata, 400 strata, 150 zones, 50 regions, and 8 countries once each month.

8.1.2 Response Requirements (Req'd by: CAS; Cat. 1;
Impl. Resp.: CAS/LACIE-ASVB)

Production of area, yield, and production reports monthly for at least eight countries requires interactive automatic data processing and computer accessible storage and retrieval of sample segment classification information. The outputs of area, yield and production data by stratum, zone, region, and country require approximately 1800 reports per month. See Appendix A for reports. It is estimated that this activity will require 8 hrs/day for 3 weeks and 16 hrs/day for 1 week each month.

8.1.3 Reliability Requirements (Req'd by: CAS; Cat. 1;
Impl. Resp.: CAS)

8.1.3.1 Data base.- The data base should be periodically recorded on magnetic tape for rapid data base recovery to prevent extensive data loss requiring more than 2 hours to reconstruct.

8.1.4 Security Requirements (Req'd by USDA; Cat. 1;
Impl. Resp.: CAS/LACIE-ASVB)

A LACIE Security Plan shall be prepared by each organization designated with implementation

responsibility. The plan for each implementing organization shall define the specific measures that will be utilized by that organization to comply with the LACIE security requirements. The LACIE security requirements shall be defined by the USDA and will be forwarded to the implementing organizations upon receipt by the LACIE Project Manager. Each implementing organization shall submit their plan for approval to the LACIE Level III Change Board within 90 days after their receipt of the USDA requirements.

8.1.5 Delivery Requirements (Req'd by: CAS; Cat. 1; Impl. Resp.: CAS/LACIE-ASVB)

The monthly crop assessment reports are transferred to USDA, Washington, D.C., by registered mail each month.

8.1.6 Quality Assurance Requirements (Req'd by: LACIE-PO; Cat. 1; Impl. Resp.: CAS/LACIE-ASVB)

A LACIE Quality Assurance Plan is being prepared, as well as a separate Quality Assurance Procedures document. These documents are being prepared under the direction of the Quality Assurance Manager.

8.1.7 Other Operational Requirements

None.

8.2 ACCURACY ASSESSMENT

8.2.1 Throughput Requirements (Req'd by: CAS; Cat. 1; Impl. Resp.: ISRRS)

The throughput requirements of the accuracy assessment are dependent upon the throughput requirements of other LACIE functional components. That is, as the number of segments processed per day or per week increases, the amount of data processing required will correlatively increase, and vice versa.

8.2.2 Response Requirements (Req'd by: CAS; Cat. 1;
Impl. Resp.: ISRRS)

The CAS will require one-day turnaround on error model runs. The CAS will also require one-day response for data retrieval from the ISRRS.

8.2.3 Reliability Requirements (Req'd by: CAS; Cat 2;
Impl. Resp.: ISRRS)

The reliability requirements (up time), of the LACIE system, for the CAS must match those required by the CAMS.

8.2.4 Security Requirements (Req'd by: CAS; Cat. 1;
Impl. Resp.: IE)

The CAS security requirements are defined in the IE/CAS ICD.

8.2.5 Delivery Requirements

The TBD delivery requirements of and for CAS data is dependent upon the security requirements of the data being delivered.

8.2.6 Quality Assurance Requirements

A LACIE Quality Assurance Plan is being prepared, as well as a separate Quality Assurance Procedures document. These documents are being prepared under the direction of the Quality Assurance Manager.

~~REEDING PAGE BLANK NOT PLANNED~~

SECTION 9.0

VERIFICATION REQUIREMENTS

9.1 CROP ASSESSMENT (Req'd by: CAS; Cat. 2; Impl. Resp.: CAS/LACIE-ASVB)

The CAS shall prepare the subsystem portion of the LACIE Verification Plan and submit it for Level III approval 90 days after the Level IV Requirements Documents are baselined. The plan shall provide a complete definition of the verification functions required to verify the operation of the CAS portion of the LACIE system. Each portion of the software package shall be verified prior to integration into the operational system.

9.2 ACCURACY ASSESSMENT

Each organization designated with implementation responsibility shall prepare a LACIE Verification Plan, including a complete definition of the verification functions proposed for verification of the portion of LACIE for which they are responsible. Each organization shall submit their plan for approval to the LACIE Level III change board within 90 days after the requirements documents are baselined. As a minimum, the verification plan shall include the following related to the CAS subsystem:

- A. Program verification. (Required by: CAS; Category 1; Implementation Responsibility: DSAD). DSAD will provide to CAS current source listings and flow diagrams of all programs implemented for LACIE.
- B. System testing. (Required by: CAS; Category 2; Implementation Responsibility: DAPTS/CAMS/YES/CAS/ISRRS). The CAS will provide a set of test data to each of the following subsystems: DAPTS, CAMS, YES, CAS, and ISRRS. Each of these subsystems will use their standard operating procedure to input this test

data into the LACIE systems. All accuracy assessment routines and procedures will then be exercised and their output results verified. The test results will be supplied to the Quality Assurance Manager.

SECTION 10.0

RESEARCH REQUIREMENTS

10.1 CROP ASSESSMENT (Req'd by: CAS; Cat. 2; Impl. Resp.: Research)

The Crop Assessment Subsystem required the following research to be conducted.

10.1.1 Pre-Phase I

Prior to Phase I, the research effort will be required to provide a plan (including schedules) to CAS for the conduct of research into improvements of sampling strategies and identification of software and hardware necessary for the implementation of these techniques.

10.1.2 Desirable Approaches

Approaches that may be considered (but not necessarily limited to) are:

10.1.2.1 Sampling.- (Required by: CAS; Category 1; Implementation Responsibility: .CIE-RTEB).

- A. Develop a standard sampling strategy for all LACIE countries, taking into account the unique characteristic of each country. Sampling strategies required by beginning of Phase IC.
- B. Develop variable sampling techniques for area estimation. A study is to be made of the feasibility of using photointerpretation data for the purpose of allocating sample segments.
- C. Define improvement of sampling strategy produced by full-frame registered data as follows:

1. Develop a model for the variance of an acreage estimate based on optimal allocations of samples without regard to present restricting (i.e., 5x6 mi size and "Landsat constraints"). Such a model should include not only a sampling variance, but also a component of degradation due to signature extension errors. Some provision for training areas should also be made (Category 2).
 2. Attempt to quantify the above by obtaining estimates of variances from work done by CAMS on the intensive sites, as well as historical data for the last ten years of wheat areas by county over the intensive test site areas (Category 2).
- D. Evaluate reduction in sampling error due to "cluster sampling"; provide means to balance increased sample error versus reduced classification error when classification error dependence upon signature extension is known; perform such balancing (Category 2).
- E. Determine the effects on the sample strategy due to the increased overlap in the Landsat coverage at high latitudes and provide a technique of correcting for any bias that results from the increased overlap. For example, the overlap at 29° N is 7.3 percent and at 50° N, it is 44.6 percent. This increases the probability of obtaining cloud free acquisition of sample segments at the higher latitudes that could bias the sampling strategy (Category 1).

10.1.2.2

Aggregation.- (Required by: CAS; Category 1;
Implementation Responsibility: LACIE-RTEB).

- A. Develop a standard aggregation scheme for all LACIE countries (based on task 10.1.2.2 above).
- B. Devise an aggregation scheme which preferentially incorporates sample segments of higher confidence in wheat area estimates. The wheat

area estimate for a given substratum (i.e, county) will, in general, be a weighted average of the historical (ratio) and the current (Landsat) estimates. If all segments in a county are lost due to clouds, then the weights would be 0 for "Landsat" and 1 for "ratio". Conversely, if all data is good, the weights might be 0 for "ratio" and 1- for "Landsat" where is to be determined (possibly zero). If some data is available but not all, an optimal system of weight determination must be developed.

- C. Determine a model for the inclusion of cloud cover in the sampling area. Cloud cover statistics will be combined with the aggregation scheme defined above to provide estimates of sampling errors introduced by cloud cover data loss.

10.1.2.3 Harvested Wheat.- (Required by: CAS; Category 2; Implementation Responsibility: LACIE-RTEB). Develop a technique for estimating the harvested wheat area from a combination of the percent harvested wheat on a given date from Landsat data with the harvest start and stop dates.

10.2 ACCURACY ASSESSMENT (Req'd by: CAS; Cat. 3; Impl. Resp.: LACIE-RTEB)

Investigation in the following general areas is required to study the error propagation in the LACIE system and to support the error analysis efforts.

- A. Effect of AI misidentification of training fields on the classification performance.
- B. Estimation of variability in the area estimate due to AI selection procedure for training fields.
- C. Effect of signature extension on classification accuracy when signature extension is considered

as a function of distance, direction, time-phase and properties of analogous areas.

- D. Relationship between the number of classes/subclasses considered, the number of channels used and the number of samples employed in regard to classification accuracy and the area estimation.
- E. Effect of data quality as a function of mis-registration, cloud cover and haze, data dropout, on the classification performance.
- F. Different area estimation techniques for the bias and the mean square error of a crop proportion estimate.
- G. Methods for obtaining crop inventory estimates when part of the data is missing due to cloud cover and haze, data dropouts, etc.
- H. Development of an error model that accounts for different types of errors likely to arise in the operation of LACIE system.

10.2.1

Error Model (Req'd by: CAS; Cat. 1; Impl. Resp.: LACIE-RTEB)

The following interim error budget has been defined to replace the existing 7-7-2 budget (7 percent area, 7 percent yield, 2 percent sampling) at a national level, with a best estimate of an error budget which will permit a 90/90 criteria for production, at harvest on a National Level. When the TRW developed error model becomes available this interim error budget will be updated:

In order for the production estimate from LACIE to be 90 percent accurate - 90 percent of the time a coefficient of variation of the production estimate CV (P) of 6 percent or less is permissible provided the estimate is unbiased. Wheat production estimate (P) is obtained from the product of the wheat area estimate (A) and yield estimate (Y), and A and Y are independent then:

$$CV(P) = ((CV(A) CV(Y))^2 + CV(A)^2 + CV(Y)^2)$$

If the allowable error is equally shared by area and yield, then:

$$\begin{aligned} CV(A) &= CV(Y) = V \\ \text{Letting } CV(P) &\leq .06 \text{ then} \\ V^4 + 2V^2 &= .0036 \end{aligned}$$

So, the allowed errors for area and yield
 $CV(A) = V \leq .0424$ now if sampling error,
 $CV(SE)$ is arbitrarily assigned to be 2 percent,
then:

$$CV(A) = (CV(SE)^2 + CV(CAMS)^2)$$

So, the initial LACIE error budget is:

$$\begin{aligned} CV(CAMS) &\leq 3.74 \text{ percent} \\ CV(SE) &\leq 2 \text{ percent} \\ CV(Y) &\leq 4.26 \text{ percent} \end{aligned}$$

If a bias in the production estimate is occurring, the errors in the estimate of yield and area which are allowed would have to be reduced. For example, for a bias of 2 percent in the production estimate, the $CV(P) = 5.4$ percent and

$$\begin{aligned} CV(CAMS) &\leq 3.25 \text{ percent} \\ CV(SE) &\leq 2 \text{ percent} \\ CV(Y) &\leq 3.82 \text{ percent} \end{aligned}$$

$$\text{The } CV(CAMS)^2 = \frac{CV(SS)^2}{n}$$

where $CV(CAMS)$ = overall CAMS CV
 $CV(SS)$ = CV on a sample basis for CAMS
 n = number of sample segments
acquired

Thus for $n = 200$

CV(SS) = 37.4 percent

These error allocations will be revised in January 1976 and August 1976, using the accuracy assessment and error model results.

SECTION 11.0

TEST AND EVALUATION REQUIREMENTS

11.1

CROP ASSESSMENT

(Req'd by: CAS; Cat. 2; Impl. Resp.: LACIE-RTEB)

The Crop Assessment Subsystem requires the following test and evaluation to be conducted:

- A. Prior to Phase IC of the project, the Test and Evaluation will be required to plan and implement test and evaluation of the procedures for calculating acreage and sensitivity of the Sampling Strategy.
- B. Prior to Phase II of the project, the Test and Evaluation will be required to plan and implement test and evaluation of the procedures for calculating production and sensitivity of the Sampling Strategy.

11.2

ACCURACY ASSESSMENT (Req'd by: CAS; Cat. as indicated; Impl. Resp.: LACIE-RTEB)

The Test and Evaluation is required to evaluate the significance of each major source of error and their interrelationships with respect to the classification performance and the area estimation. The following evaluations are required to be carried out:

- A. Significance of error sources
 1. Evaluate the effect of misregistration on the AI and classification performances (Category 2).
 2. Evaluate the effect on classification accuracy due to signature extension when considered as a function of distance, direction, time-phase and properties of analogous areas (Category 2).

3. Evaluate the significance of AI selection for the training fields and the number of these fields on the area estimation and the classification accuracy (Category 2).
 4. Evaluate the effect of AI misidentification of fields on the classification performance (Category 2).
 5. Evaluate the effect of classification performance on the area estimation, considering the classification performance as a function of each of the following factors (Category 2):
 - a. number of classes/subclasses
 - b. number of features
 - c. number of samples
 - d. threshold limits
 - e. a priori probabilities
 - f. clustering algorithms used or not used
 6. Evaluate the sampling scheme for the sampling properties of the area estimate (Category 2).
- B. Interrelationship between different types of performances. Evaluate the following relationships with respect to the classification performance and the area estimation:
1. The relationship between the number of features used and the number of classes/subclasses considered (Category 2).
 2. The relationship between the number of features used and the number of samples employed (Category 2).
 3. The relationship between a signature extension factor and a classification technique factor. These factor types

are listed in 11.2 A.2 and 11.2 A.5
(Category 3).

4. The relationship between the AI performance and the ADP analyst performance (Category 3).

- C. Test and evaluate. Test and Evaluation are required to assess all error models developed and/or utilized by the CAS to assess performance of the LACIE system (Category 1).

APPENDIX A

SAMPLE SEGMENT ALLOCATION PROCEDURE FOR U.S.A.

RATIONALE AND PROCEDURE FOR SAMPLE SEGMENT
ALLOCATION AND LOCATION FOR THE U.S.A.

A1.0 SAMPLE SEGMENT ALLOCATION

All of the following calculations were made using the USDA wheat acreage data for the year of 1969.

- A. Calculate the proportion of wheat acres for each county:

$$P_k = \text{Acres of wheat in } k^{\text{th}} \text{ county} / \text{total acres in } k^{\text{th}} \text{ county}, \quad P_k = Aw_k / At_k$$

- B. Calculate the standard deviation of the proportion of wheat for each county:

$$\sigma_k = \sqrt{P_k(1-P_k)}$$

- C. Calculate the threshold value based on the total acres and standard deviation for each county:

$$t^* = At_k \times \sigma_k$$

- D. Calculate the sum of all the threshold values for the country:

$$t^* = \sum_k^N t^*_k$$

where N is the total number of counties in the country.

- E. Based on 638 sample segments for the U.S.A. (n), calculate the proportional number of sample segments allotted to each county:

$$\frac{(n_k)^1}{638} = \frac{t^*_k}{t^*} \quad \text{or} \quad (n_k) = \frac{t^*_k \times 638}{t^*}$$

(n_k) = unrounded calculation

- F. Repeat steps 4 and 5 for each Crop Reporting District (CRD) in the country:

$$F.1 \quad t^*_j = \sum_{k=1}^L t^*_k$$

where L is the number of counties in the j^{th} CRD.

$$F.2 \quad (n_j) = \frac{t^*_j \times 638}{t^*}$$

(n_j) = unrounded calculation

n_j = rounded integer value giving the number of segments to be allocated to the j^{th} CRD in the U.S.A.

¹Circle indicates unrounded calculations.

- G. Stratified Sample Segment.- All counties that have an $\bar{n}_k \geq 0.500$ will have at least one sample segment; the actual number of segments is the rounded value of $\bar{n}_k$ or n_k .
- H. No Sample Segments.- All counties that have an $\bar{n}_k < 0.100$ will receive no sample segments.
- I. Probability Proportional to Size (PPS) Sample Segment.- All counties that have an $\bar{n}_k < 0.500$ and $\bar{n}_k > 0.100$ will have a probability of having a sample segment that is proportional to the magnitude (numerical size) of their $\bar{n}_k$ with respect to the value of n_j of the CRD in which the county is. Example:

$$n_j \text{ for CRD} = 5$$

This means that the CRD will have five sample segments.

<u>County</u>	<u>$\bar{n}_k$</u>	<u>Result</u>
1	0.08	no sample
2	1.63	two samples (stratified)
3	1.10	one sample (stratified)
4	0.30	PPS
5	0.40	PPS
6	0.40	PPS
7	0.33	PPS

} two samples are available

Normalize, in a linear distance, the $\bar{n}_k$ for counties 4, 5, 6, and 7 to a total of two for two samples:

County - 4 =	0.0000	0.4196
County - 5 =	0.4197	0.9790
County - 6 =	0.9791	1.5385
County - 7 =	1.5386	2.000

Select a random number between zero and one - 0.3715; this falls in the region of county - 4, therefore, it receives a sample segment. To determine the second sample segment county, add one to the random number = 1.3715, which lies in the region of county - 6.

A2.0

SAMPLE SEGMENT LOCATION

A. Equipment required

- (1) Nine Operational Navigation Charts (ONC's) covering the U.S.
- (2) Dot matrix, to the scale of the ONC's, spaced six nautical miles in the east-west direction, and spaced five nautical miles in the north-south direction
- (3) A 100 x 100 nautical mile overlay, to the scale of the ONC's, with a center point surrounded by a 6 x 5 nautical mile rectangle, surrounded by a 12 x 10.5 nautical mile rectangle. The boundaries of this larger rectangle are drawn out to the edges of the 100 x 100 nautical mile Landsat (frame boundaries). Running north-south, 10 nautical miles inside the east and west boundaries of the 100 x 100 nautical mile limits making a rectangle 80 nautical miles in the east-west direction, and 100 nautical miles in the north-south direction
- (4) Lists of random numbers covering the ranges of 1 to 10, 1 to 20, 1 to 30, etc., up to 1 to 300
- (5) Data form indicating: State, CRD, county, if more than one segment in the county indicate which segment is being selected; and space to enter: number of segments (5 x 6 nautical mile rectangles) in the county, number

of segments in nonagricultural areas of the county, number of segments in the agricultural areas of the county, Landsat-1 track number passing through county, latitude and longitude of center of selected sample segment number, spring/winter wheat sample segment

B. Procedure

- (1) Fasten the dot grid to an ONC
- (2) From the data form, select a county; locate it and draw its boundaries on the ONC
- (3) Count the number of segments in the county - a segment is in the county if its center point is in the county - and enter on data form. If a center point of a segment falls exactly on a county boundary, flip a coin. If it comes up heads, place the segment in the county being worked on, if it comes up tails, place the segment in the adjacent county
- (4) Count the number of segments that fall entirely in nonagricultural areas and enter on form
- (5) Subtract (d) from (c) and record on data form
- (6) Starting with most northwest segment in (e), number all segments determined in (e)
- (7) Select lowest random number table that includes the last number entered in step (f)
- (8) Use first unused random number from table; this determines which segment is the sample segment

C. Constraints

- (1) Place the center point of the 100 x 100 nautical mile overlay on the center point of the sample segment
- (2) Is there another sample segment within the 12 x 10.5 nautical mile rectangle?
 - *Yes - discard newest sample segment and go back to step 2(h)
 - *No - proceed
- (3) Are there more than four sample segments between the extended boundaries of the 12 x 10.5 nautical mile rectangle running in the east-west direction?
 - *Yes - discard newest sample segment and go back to step 2(h)
 - *No - proceed
- (4) If there are any other sample segments falling within the east-west boundaries of the 12 x 10.5 nautical mile rectangle, move the center point of the overlay to center points of these segments to determine if the new sample segment causes there to be more than four sample segments within these boundaries -
 - *Yes - discard newest sample segment and go back to step 2(h)
 - *No - proceed
- (5) With the center point of the overlay back on the center point of the newest sample segment; are there more than eight sample segments between the extended boundaries of the 12 x 10.5 nautical mile rectangle running in the north-south direction?

*Yes - discard newest sample segment
and go back to 2(h)

*No - proceed

- (6) If there are any other sample segments falling within the north-south extended boundaries of the 12 x 10.5 nautical mile rectangle, move the center point of the overlay to the center points of these segments to determine if the new sample segment causes more than eight sample segments to be within these boundaries -

*Yes - discard newest sample segment
and go back to step 2(h)

*No - proceed

- (7) Determine if segment is a spring or winter wheat type segment - whichever occurs for more than 50 percent of the total wheat acreage
- (8) Record remaining data for this sample segment
- (9) Return to step 2(b)

REJECTING PAGE BLANK NOT
APPENDIX B

SAMPLING AND ESTIMATION PLAN

B1.0 SCOPE

This appendix describes the sampling and aggregation requirements for LACIE Phase III and subsequent activities. These requirements include a description of the sampling scheme as well as procedures for determining estimates of wheat area, yield, and production and their associated variances.

B2.0 SAMPLE SELECTION PROCEDURE

The sampling strategy to be used in LACIE will depend on whether a country has detailed historical data (U.S., Canada, Australia) or whether it only has data on one level smaller than the country itself (U.S.S.R., China, Argentina, Brazil, India). In the latter case, a standard stratified sampling procedure will be employed, while in the first situation the sampling strategy will consist of a two-stage stratified probability sample in which "substrata" (smallest political area for which yields, areas, and crop calendars are published) represent the primary sampling units.* The 5 by 6 nautical mile segments are the secondary units.

Sample segments will be allocated to the substrata based on weights, w_{ijk} , (see appendix A) which are a function of (a) agricultural area in the substratum, (b) within-substratum standard deviation of small grain area from segment to segment, (c) classification error variance, (d) substratum yield, and (e) substratum yield prediction error. Depending upon the w_{ijk} , substrata are designated as Group I (high sampling rate), Group II (low sampling rate) and Group III (not sampled). (In the U.S., where substrata are counties, the range of variation of the

* This allocation procedure was used in U.S. Great Plains for Phase III.

w_{ijk} is such that in order to meet the 90-90 criterion in the Great Plains, anywhere from 0 to 5 segments to a county are allocated. In foreign areas where the strata are larger, several times that number may be assigned.)

The following sections (B2.1 and B2.2) describe the allocation process.

B2.1 PRIMARY SAMPLING

A. DETERMINATION OF TOTAL SAMPLE SIZE

The total sample size allocated for a country should be such that the LACIE production estimate satisfies the "90-90" criterion allowing for errors due to (a) sampling, (b) classification, (c) yield prediction, and (d) loss of data.

Using the best available a priori estimates of the magnitude of the above errors, one can approximate the mean-squared prediction error for production as a function of the total sample size, n , (see appendix A).

$$PE(\hat{P}) = f(n; \delta) \quad (B2.1-1)$$

where $\hat{P}$ is the LACIE production estimate, δ is a set of error parameters, and $PE(\hat{P})$ is the root mean squared prediction error of $\hat{P}$.

In order to meet the "90-90" criterion, assuming $\hat{P}$ is normally distributed with expectation P , we would like $\frac{PE(\hat{P})}{P} \leq .06$. While it is not unreasonable to assume $\hat{P}$ is normally distributed, there may be some bias in $\hat{P}$. Therefore, we arrive at a sample size by setting $\frac{PE(\hat{P})}{P} = \alpha$ where α is smaller than .06. Then equation (B2.1-1) can be solved for the

* This choice was based on an acquisition rate of approximately 80 percent.

total sample size, n . (Current plans call for $\alpha = .05$ in the U.S. This results in a sample size of 601 for the Great Plains.)*

B. DESIGNATION OF GROUP III

As denoted in appendix A, let P_{ijk} denote the historical production (Primary Epoch Year) for the k th substratum, j th stratum and i th zone.

Let

$$P_0 = \max \left\{ P \mid \sum_{ijk \in S(P)} P_{ijk} < .025 P_T \right\}$$

where

$$S(P) = \left\{ P_{ijk} \mid P_{ijk} < P \right\} \text{ and } P_T = \sum_{ijk} P_{ijk}$$

(use ordering of P_{ijk} 's to obtain the set $S(P)$ and hence P_0).

Then all substrata with $P_{ijk} < P_0$ are excluded from sampling and designated "Group III." No segments will be allocated to these substrata and the ratio estimation will be used to make their area estimates.

C. DESIGNATION OF GROUPS I AND II

For all the substrata which are not Group III, let

$$w_{ijk}^* = \frac{w_{ijk}}{\sum'_{ijk} w_{ijk}} n$$

where Σ' denotes summation only over the nonexcluded substrata; then:

- a. All substrata with $w_{ijk}^* \geq 1.0$ are "Group I"
- b. All substrata with $w_{ijk}^* < 1.0$ are "Group II"

D. COMPUTATION OF SAMPLE SIZES FOR SUBSTRATA

Specific to this discussion only, an "element" shall mean either (a) a Group I substratum or (b) the union of all Group II substrata within a stratum. Let t_{ijl} represent the "size" of the l th element in the j th stratum and i th zone; then $t_{ijl} = w_{ijk}^*$ if the element is the k th Group I substratum or

$$t_{ijl} = \sum_{k \in II} w_{ijk}^*$$

where " $\sum_{k \in II}$ " denotes the summation over Group II substrata in the j th stratum and i th zone.

For each zone i , do the following:

- a. Define $n_i = \text{rounded } \sum_j \sum_k t_{ijk}$.
- b. Let n_{ijk}^0 be the provisional allotment where $n_{ijk}^0 = \text{integer part of } t_{ijk}$. Also let

$$r_{ijk} = t_{ijk} - n_{ijk}^0$$

and

$$r = \left(n_i - \sum_j \sum_k n_{ijk}^0 \right)$$

- c. Assign one additional sample to each of r elements with probabilities proportional to r_{ijk} ; i.e., let

$$n_{ijk} = n_{ijk}^0 + \delta_{jk}$$

where $\delta_{jk} = 1$ if the jk th element receives an extra sample segment; $\delta_{jk} = 0$ otherwise. If $n_{ijk} = 0$ for a Group II element, then all substrata in the element are designated as Group III.

- d. For each Group II element assigned a total of n_{ijk} sample segments, choose the substrata which receive a segment with probabilities proportional to their wheat area in some "Epoch Year." Once these substrata are chosen, the probabilities of selection, " π_{jk} ", are fixed and used in the Group II acreage estimate (see section B3.2) even though the designated "Epoch Year" may change.

The Hartley-Rao procedure¹ will be utilized to perform the PPS sampling. The details of this procedure are documented in appendix C for completeness.

B2.2 SECONDARY SAMPLING

Once the number of segments in each substratum has been determined from B2.1, the secondary units (5 by 6 segments) are selected at random within the agricultural area of the substratum, subject to certain engineering constraints. See appendix A for details of this selection process.

B3.0 AREA AGGREGATION

For zones* which do not contain at least three processed sample segments do not compute an area estimate. In all other cases, wheat area estimates at the segment level will be aggregated according to whether

¹ c.f. Cochran.

* Strata in the case of countries without detailed historical data.

they came from a Group I or Group II substratum. Estimates for substrata in Group III are obtained by the ratioing procedure described in section B3.3. The general aggregation procedure is as follows:

- a. For all substrata with $w_{ijk}^* \geq 1.0$ (Group I), and for which data exists, compute a stratified random sampling estimator of the total wheat area (see section 3.1). All substrata (in Group I) for which no samples were available because of cloud cover are placed in Group III. If data subsequently becomes available, a substratum is reinstated as Group I.
- b. For all substrata with $w_{ijk}^* < 1.0$ (Group II) compute a PPS estimator of the wheat area. (See section 3.2.) If an entire set of Group II substrata within a CRD has no available segments due to cloud cover, the group of substrata is treated as Group III until data become available.
- c. For all remaining substrata, plus those for which no samples are available due to cloud cover, (Group III), compute a ratio estimate of the wheat area (see section 3.3).
- d. The final wheat area estimate shall be the sum of the estimates for Groups I, II, and III.

B3.1 GROUP I AGGREGATION

The wheat area estimate is calculated by the following equations:

$$A_{1jk} = \frac{N_{1jk} R_{1jk}}{M_{1jk}} \sum_{i=1}^{M_{1jk}} A_{1jki} \quad (B3.1-1)$$

at the substratum level and

$$A_{ij} = \sum_{k=1}^{L_{ij}} A_{1jk} \quad (B3.1-2)$$

at the stratum level where

A_{1j} = The combined wheat area estimate for the Group I substrata in the j th stratum.

L_{1j} = The number of Group I substrata for which classification data are available in the j th stratum.

A_{1jk} = The wheat area estimate of the k th Group I substratum in the j th stratum.

N_{1jk} = The number of agricultural segments (after exclusion) in the k th substratum.

M_{1jk} = The number of sample segments for which classification data are available in the k th substratum.

R_{1jk} = The ratio of the true substratum area to the gross pseudo substratum (before exclusion) area

A_{1jki} = The estimated wheat area for the i th sample segment in the k th substratum. (See Appendix E for determination of wheat proportion in a segment.)

B3.2 PROBABILITY-PROPORTIONAL-TO-SIZE (GROUP II) AGGREGATION

A PPS area estimate is computed for the group of substrata within a stratum that are designated as Group II. The wheat area estimate is calculated using the following equation.

$$A_{2j} = \sum_{k=1}^{M_{2j}} R_{2jk} \frac{A_{2jk}}{\pi_{2jk}} N_{2jk} \quad (B3.2-1)$$

A_{2jk} = The wheat area estimate of the sample segment belonging to the k th substratum in the j th stratum.

M_{2j} = The number of sample segments in the PPS group for which classification data are available.

N_{2jk} and R_{2jk} are defined for Group II similar to N_{1jk} and R_{1jk} for Group I.

The π_{2jk} are proportional to the historical wheat area in the k th substratum, j th stratum for the designated Primary Epoch Year at the time of sample selection (see section B2.1) and satisfy $\sum_k \pi_{2jk} = M_{2j}$. Specifi-

cally, $\pi_{2jk} = \frac{M_{2j} W_{2jk}}{W_{2j}}$ where W_{2jk} = the primary epoch year harvested wheat area in the k th Group II substrata in the j th stratum and $W_{2j} = \sum_k W_{2jk}$.

B3.3 GROUP III AGGREGATION

The wheat area for the Group III substrata will be estimated as follows:

If there are at least three sample segments processed in the j th stratum, the Group III area estimate will be

$$A_{3j} = \left(\frac{A_{1j} + A_{2j}}{W_{1j} + W_{2j}} \right) W_{3j}$$

where W_{1j} = the sum of the primary epoch year harvested wheat area in the Group I area in the j th stratum and W_{3j} = the sum of the primary epoch year harvested wheat area in the Group III area in the j th stratum, and W_{2j} is as defined in Section B3.2.

If there are less than three sample segments processed in the j th stratum, the Group III area estimate will be as follows:

$$A_{3j} = \left(\frac{\sum_{j \in I} A_{1j} + \sum_{j \in II} A_{2j}}{\sum_{j \in I} W_{1j} + \sum_{j \in II} W_{2j}} \right) W_{3j}$$

where " $\sum_{j \in I}$ " denotes summation over the Group I areas

and " $\sum_{j \in II}$ " denotes summation over the Group II areas
for which classification data are available in the zone.

B3.4 COMBINED STRATUM ESTIMATE

The wheat area estimate of each stratum will be computed as the sum of the Group I, Group II, and Group III component estimates which comprise the stratum,

$$A_j = A_{1j} + A_{2j} + A_{3j} \quad (B3.3-1)$$

If the stratum does not contain a Group III element then the sum of the Group I and Group II components define the strata estimate.

$$A_j = A_{1j} + A_{2j} \quad (B3.3-2)$$

B4.0 AREA VARIANCE ESTIMATION

In countries which lack detailed historical data no variance estimates are made for strata containing less

~~scribble~~
C2

than 3 processed segments. For other strata in such countries, the strata variance estimates are obtained using equations B4.1-1 and B4.3-1. In countries with detailed historical data the procedure described below in sections B4.1 to B4.4 will be followed.

B4.1 GROUP I SUBSTRATUM VARIANCE ESTIMATE

The variance of the wheat area estimate for the Group I substrata is computed by:

$$v_{1j} = \sum_{k=1}^{L_{1j}} R_{1jk}^2 \frac{N_{1jk}^2}{M_{1jk}} S_{1jk}^2 \quad (B4.1-1)$$

where L_{1j} , R_{1jk} , N_{1jk} , and M_{1jk} are as defined in section B3.1, and S_{1jk}^2 is the within substratum variance to be computed according to section B4.3.

B4.2 VARIANCE OF GROUP II/PPS ESTIMATE

The variance of the wheat area estimate for the probability proportional to size (Group II) substrata is computed by:

$$v_{2j} = \sum_{k=1}^{L_{2j}} \frac{N_{2jk}^2}{\pi_{2jk}} R_{2jk}^2 S_{2jk}^2 + \sum_{k=1}^{L_{2j}-1} \sum_{i=k+1}^{L_{2j}} (\pi_{2jk} \pi_{2ji} - \pi_{2jki}) \left(\frac{Y_{2jk}}{\pi_{2jk}} - \frac{Y_{2ji}}{\pi_{2ji}} \right)^2 \quad (B4.1-2)$$

where L_{2j} = number of PPS substrata in j th stratum and π_{2jk} , L_{2j} , M_{2j} , N_{2jk} , and R_{2jk} are as defined in section B3.2, and S_{2jk}^2 is the within substratum variance to be computed according to section B4.3, and for $i \neq k$, π_{2jki} is computed using the following formula:

$$\begin{aligned} \pi_{2jki} = & \frac{M_{2j}^{-1}}{M_{2j}} \pi_{2jk} \pi_{2ji} \\ & + \frac{M_{2j}^{-1}}{M_{2j}^2} \left(\pi_{2jk}^2 \pi_{2ji} + \pi_{2jk} \pi_{2ji}^2 \right) \\ & - \frac{M_{2j}^{-1}}{M_{2j}^3} \left(\pi_{2jk} \pi_{2ji} \right) \left(\sum_{\alpha=1}^{L_{2j}} \pi_{2j\alpha}^2 \right) \\ & + \frac{2(M_{2j}^{-1})}{M_{2j}^3} \left(\pi_{2jk}^3 \pi_{2ji} + \pi_{2ji}^3 + \pi_{2jk}^2 \pi_{2ji}^2 \right) \\ & - \frac{3(M_{2j}^{-1})}{M_{2j}^4} \left(\pi_{2jk}^2 \pi_{2ji} + \pi_{2jk} \pi_{2ji}^2 \right) \left(\sum_{\alpha=1}^{L_{2j}} \pi_{2j\alpha}^2 \right) \\ & + \frac{3(M_{2j}^{-1})}{M_{2j}^5} \left(\pi_{2jk} \pi_{2ji} \right) \left(\sum_{\alpha=1}^{L_{2j}} \pi_{2j\alpha}^2 \right)^2 \\ & - \frac{2(M_{2j}^{-1})}{M_{2j}^4} \left(\pi_{2jk} \pi_{2ji} \right) \left(\sum_{\alpha=1}^{L_{2j}} \pi_{2j\alpha}^3 \right) \end{aligned}$$

Also,

Y_{2jk} = Secondary epoch year harvested wheat area in kth Group II county in the jth stratum.

Y_{2ji} = Secondary epoch year harvested wheat area in ith Group II county in the jth stratum.

NOTE: If the secondary epoch year data are not available then set the second term of the variance of A_{2j} equal to $\frac{T_1}{e\sqrt{2\pi}}$, where T_1 is the first term of V_{2j} given in equation B4.1-2.

B4.3 COMPUTATION OF S_{1jk}^2 AND S_{2jk}^2 (WITHIN SUBSTRATUM/STRATUM VARIANCE)

If the country lacks detailed historical data and hence has large strata, only the variance S_{1jk}^2 needs to be computed and is given by

$$S_{1jk}^2 = \sum_{i=1}^{M_{1jk}} (A_{1jki} - \bar{A}_{1jk})^2 / (M_{1jk} - 1) \quad (B4.3-1)$$

If the country has detailed historical data, segments within a zone are grouped into collections according to their corresponding a priori within substratum standard deviations, θ_k , used in the original allocation (see appendix A). These collections should satisfy the following conditions:

- a. No collection should contain less than 8 segments unless there are less than 8 segments available for the whole zone.
- b. All segments in the same substratum shall be in the same collection.

- c. The number of collections c shall be given initially by:

$c = 1$ provided $NS < 16$

$c = 2$ provided $16 \leq NS < 24$

$c = 3$ provided $NS \geq 24$

where NS = the number of available segments in the zone. If (a) and (b) above cannot be satisfied when $NS \geq 16$, reduce the value of c by one.

- d. If $c > 1$, the collections should maximize the ratio of the between to within variance of the θ_k ; i.e., let

$$F = \frac{\sum_{r=1}^c \sum_{j=1}^{d_r} (\bar{\theta}_{r\cdot} - \bar{\theta}_{..})^2 / (c - 1)}{\sum_{r=1}^c \sum_{j=1}^{d_r} (\theta_{rj} - \bar{\theta}_{r\cdot})^2 / (\sum_r d_r - c)}$$

(B4.3-2)

where θ_{rj} = a priori small grains standard deviation (see appendix A) associated with the substratum containing the j th segment in the r th collection (note that segments in the same substratum have duplicate θ_{rj} values) and d_r = number of segments in r th collection. Then the partitioning of the $\sum_r d_r$ segments into collections should be such that F is maximized subject to the constraints specified in (a) and (b) above. If $F < 1$ reduce c by 1.

Let $\hat{p}_{rj}$ be the CAMS wheat proportion estimate for the j th segment of the r th collection and let x_{rj} be the Epoch year wheat proportion for the substratum containing that segment. Then for each of c collections, regress the p_{rj} on the x_{rj} and let S_r^2 denote the residual mean square, i.e.,

$$S_r^2 = (\text{segment area})^2 \sum_{j=1}^{d_r} (\hat{p}_{rj} - p_{rj})^2 / (d_r - 2)$$

where $p_{rj} = a + bx_{rj}$ the predicted value using the regression equation)

$$\bar{a} = \bar{\hat{p}}_{r\cdot} - b\bar{x}_{r\cdot}$$

$$b = \frac{\sum_j \hat{p}_{rj} (x_{rj} - \bar{x}_{r\cdot})}{\sum_j (x_{rj} - \bar{x}_{r\cdot})^2} \quad (\text{B4.3-3})$$

Since by constraint (b), every substratum is associated with one and only one collection, assign the variance S_r^2 to every substratum whose segments are in the r th collection (whether the substratum is Group I or II), i.e., S_{1jk}^2 (or S_{2jk}^2) = S_r^2 if the segments in the k th substratum, j th stratum belong to the r th collection.

B4.4

OVERALL VARIANCE ESTIMATE

In a stratum having at least three sample segments processed, the estimate of variance for the j th CRD area estimate is obtained by

$$v_j = \left(1 + \frac{w_{3j}}{w_{1j} + w_{2j}} \right)^2 (v_{1j} + v_{2j}) \quad (\text{B4.4-1})$$

where w_{1j} , w_{2j} , and w_{3j} are defined in sections B3.2 and B3.3.

If the j th stratum has at least one but less than three segments processed, the estimate of variance for the stratum area estimate is given by

$$v_j = \left(1 + \frac{2w_{3j}}{\sum_{i \in \mathcal{I}} (w_{1i} + w_{2i})} \right) (v_{1j} + v_{2j}) + \left(\frac{w_{3j}}{\sum_{i \in \mathcal{I}} (w_{1i} + w_{2i})} \right)^2 \sum_{i \in \mathcal{I}} (v_{1i} + v_{2i}) \quad (\text{B4.4-2})$$

where $\mathcal{I}$ is the set of indices i , such that the i th stratum has at least one processed segment.

If the j th stratum has no segment processed, the variance estimate for the strata area estimate is given by

$$v_j = \left(\frac{w_{3j}}{(w_{1i} + w_{2i})} \right)^2 \sum_{i \in \mathcal{L}} (v_{1i} + v_{2i}) \quad (B4.4-3)$$

Remark: There are three categories into which the Group III data may fall, each producing a uniquely determined variance estimate for the stratum. The first category is Group III data in the stratum which was estimated from Group I and Group II data within its own stratum. The second category is Group III data in the stratum which was estimated from all the Group I and Group II data within its zone as well as the Group I and Group II data within its own stratum. The third category consists of Group III data that was estimated from all the Group I and Group II data within its zone, but not from Group I or Group II data within its stratum.

B5.0 HIGHER LEVEL ESTIMATES

B5.1 AREA AGGREGATION TO THE ZONE, REGION, AND COUNTRY LEVELS

Wheat area estimates for a zone, region, and country are obtained by adding estimates for the strata included in the zone, region, and country.

B5.1.1 Area estimate at the zone level:

$$A_{rz} = \sum_{j=1}^S A_{rzj} \quad (B5.1-1)$$

Where,

S = Number of strata in the zth zone, rth region
in the country

A_{rzj} = Area estimate of the jth strata, zth zone, rth
region in the country

B5.1.2 Area estimate at the region level:

$$A_r = \sum_{z=1}^R A_{rz} \quad (B5.1-2)$$

Where,

R = Number of zones in the rth region in the country

B5.1.3 Area estimate at the country level:

$$A = \sum_{r=1}^K A_r \quad (B5.1-3)$$

Where,

K = Number of regions in the country

B5.2 VARIANCE AGGREGATION TO THE ZONE, REGION, AND COUNTRY
LEVELS

For a zone in a country lacking detailed historical
data the variance estimate is obtained by aggregating
the strata variances in the zone,

$$V_z = \sum_j V_{zj}$$

where V_{zj} is the variance of the j th stratum in the z th zone. Otherwise, use equation B5.2-1.

B5.2.1 An estimate of the variance at the zone level is given by

$$V_z = \sum_{j \in \mathcal{L}} \left(1 + \frac{W_{3j}}{W_{1j} + W_{2j}} + \frac{\sum_{i \in \mathcal{M}} W_{3i} + \sum_{i \in \mathcal{N}} W_{3i}}{\sum_{i \in \mathcal{L}} (W_{1i} + W_{2i})} \right)^2 (V_{1j} + V_{2j})$$

$$+ \left(1 + \frac{\sum_{i \in \mathcal{M}} W_{3i} + \sum_{i \in \mathcal{N}} W_{3i}}{\sum_{i \in \mathcal{L}} (W_{1i} + W_{2i})} \right)^2 \sum_{j \in \mathcal{M}} (V_{1j} + V_{2j})$$

(B5.2-1)

where $\mathcal{L}$ is the set of indices associated with strata in the zone having at least one processed segment, $\mathcal{M}$ is the set of indices associated with strata in the zone which have at least one but less than three segments processed, $\mathcal{L}$ is the set of indices associated with strata having at least three segments processed and $\mathcal{N}$ is the set of indices associated with strata containing no processed segments.

No variance estimation is required for zones which do not contain at least three processed sample segments.

B5.2.2 The variance estimate at the region level is

$$V_r = \sum_{z=1}^R V_{rz} \quad (B5.2-2)$$

B5.2.3 The variance estimate at the country level is

$$V = \sum_{r=1}^K v_r \quad (B5.2-3)$$

B6.0 SPRING AND WINTER WHEAT

B6.1 AREA AND VARIANCE ESTIMATION

In a mixed wheat area, separate area estimates will be made for the winter wheat and the spring wheat using the aggregation procedure described in sections 3.0 through 5.2 with inputs (CAMS/CAS as well as historical) of winter wheat for a winter wheat aggregation and those of spring wheat for a spring wheat aggregation except that the π_{wjk} in sections B3.2 and B4.2 are unchanged; i.e., they are based upon the historical total wheat area for the Group II sub-strata. This will provide the spring wheat and the winter wheat area estimates and their respective variance estimates at the strata and zone levels. The procedures in sections 5.1 and 5.2, are used to obtain separate winter wheat and spring wheat area estimates and the corresponding variance estimates at the regional and country levels.

B6.2 TOTAL WHEAT: AREA AND VARIANCE ESTIMATION

The total wheat area estimate in a mixed wheat area will be computed by aggregating the winter wheat and the spring wheat area estimates for the area, that is, if A_w and A_s denote the winter and spring wheat

area estimates, respectively, the total wheat area estimate, Λ_t , is given by the equation:

$$\Lambda_t = \Lambda_w + \Lambda_s$$

This will be done at the strata and higher levels.

The variance estimates for the total wheat at the strata and zone levels will be those obtained from the total wheat aggregation to be made with inputs of total wheat by CAS for the segments and historical data. For any higher level the procedure is the same as described in section 5.2.

B7.0

STANDARD STATISTICS

The standard statistics for area estimation and aggregation are composed of standard deviation (σ), coefficient of variation (CV); 90 percent confidence limits (CL), and probability of 10 percent relative error (Ω). If A_s and V_s are the area and variance estimates at some level s , statistics will be calculated as follows.

$$a. \quad \sigma_s = \sqrt{V_s} \quad (B7.0-1)$$

$$b. \quad CV_s = \frac{\sigma_s}{A_s} \quad (B7.0-2)$$

$$c. \quad CL_s^{(+)} = A_s + (1.645) \sigma_s \quad (B7.0-3a)$$

$$CL_s^{(-)} = A_s - (1.645) \sigma_s \quad (B7.0-3b)$$

$$d. \quad \Omega_s = 2 P(x) - 1 \quad (B7.0-4)$$

$$\text{where } P(x) = \left[1 - Z(x) \left(a_1 t + a_2 t^2 + a_3 t^3 \right) \right]$$

$$\text{with } t = \frac{1}{1 + px}$$

$$x = \frac{.10}{CV_s}$$

$$p = .3327$$

$$a_1 = .4362$$

$$a_2 = -.1202$$

$$a_3 = .9373$$

$$\text{and } z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

B8.0 PRODUCTION ESTIMATION

Production estimates are not generally made at the stratum level. Instead, pseudo zones are created in a zone if it is covered by more than one yield model. A pseudo zone is obtained from the intersection of a yield stratum with area strata in a zone. An estimate of the production in a pseudo zone is obtained by the product of its area estimate times its yield prediction, and these estimates are aggregated to predict zone and higher level production. In mixed areas this will be done for each crop-type (winter and spring wheat). To obtain the total production, the separate production for each crop-type will be aggregated.

B8.1 PREDICTION FOR A ZONE

Suppose the zone consists of H pseudo zones, $G_1, G_2, \dots, G_H$ with acreage estimates $A_{z1}, \dots, A_{zH}$ and yield predictions $Y_{z1}, Y_{z2}, \dots, Y_{zH}$, respectively. Then the zone production estimate is given by

$$P_z = \sum_{i=1}^H A_{zi} Y_{zi} \quad (\text{B8.1-1})$$

B8.2 PREDICTION ERROR ESTIMATE FOR A ZONE

An estimate of the squared prediction error of P_z is given by

$$S_z^2 = \sum_{i=1}^H \left(v_{zi} Y_{zi}^2 + u_{zi} A_{zi}^2 - v_{zi} u_{zi} \right) + 2 \sum_{i=2}^H \sum_{i=1}^{i-1} Y_{zi} Y_{zi'} \left(\sum_{j \in G_i} \sum_{k \in G_{i'}} v_{jk} \right) \quad (\text{B8.2-1})$$

where u_{zi} is the estimated squared prediction error of the yield for the i th pseudo zone and v_{zi} is the area variance estimate for the i th pseudo zone and is given by

$$v_{zi} = \sum_{j \in G_i} v_j + \sum_{j \in G_i} \sum_{\substack{k \in G_i \\ j \neq k}} v_{jk} \quad (\text{B8.2-2})$$

here v_j is the area variance estimate for j th stratum and is obtained according to the formula in section B4.4. v_{jk} in equations (B8.2-1) and (B8.2-2) is the estimated covariance between area estimates for the j th and k th strata in the zone and is obtained according to the following formulation:

$v_{jk} = 0$ when each of j th and k th strata has at least 3 segments processed

$$v_{jk} = \frac{w_{3k}(w_{1j} + w_{2j} + w_{3j})(v_{1j} + v_{2j})}{(w_{1j} + w_{2j}) \sum_{i \in \mathcal{I}} (w_{1i} + w_{2i})}$$

when j th strata has at least 3 segments processed and k th strata has either no segment or less than 3 segments processed.

$$v_{jk} = \frac{w_{3k}(v_{1j} + v_{2j}) + w_{3j}(v_{1k} + v_{2k})}{\sum_{i \in \mathcal{I}} (w_{1i} + w_{2i})} + \frac{w_{3j}w_{3k} \sum_{i \in \mathcal{I}} (v_{1j} + v_{2j})}{\left[\sum_{i \in \mathcal{I}} (w_{1i} + w_{2i}) \right]^2}$$

when each of j th and k th strata has less than 3 but at least one segment processed.

$$v_{jk} = \frac{w_{3k}(v_{1j} + v_{2j})}{\sum_{i \in \mathcal{I}} (w_{1i} + w_{2i})} + \frac{w_{3j}w_{3k} \sum_{i \in \mathcal{I}} (v_{1i} + v_{2i})}{\left[\sum_{i \in \mathcal{I}} (w_{1i} + w_{2i}) \right]^2}$$

when j th strata has less than 3 but at least one segment processed and k th strata with no segment processed.

$$v_{jk} = \frac{w_{3j} w_{3k}}{\left[\sum_{i \in \mathcal{L}} (w_{1i} + w_{2i}) \right]^2} \sum_{i \in \mathcal{L}} (v_{1i} + v_{2i})$$

when each of j th and k th strata has no segment processed.

The quantities v_{1j} , v_{2j} , w_{1j} , w_{2j} , w_{3j} and the set $\mathcal{L}$ are as defined earlier in sections B3.0 through 5.2.

In a zone with both spring and winter wheat, pseudo zones will be of 3 types, pure spring, pure winter and mixed. These pseudo zones will be collections of strata to be specified by the MWPZDC (Mixed Wheat Pseudo Zone Definition Committee). The variance estimate for total wheat in the zone is obtained from (B8.2-1) where Y_{zi} and U_{zi} are as previously defined for the pure pseudo zones (either winter or spring).

For a mixed wheat pseudo zone, a combined yield prediction and its squared error, i.e., Y_{zi} , and U_{zi} , are to be evaluated as follows:

$$Y_{zi} = \frac{A_{wo} Y_w + A_{so} Y_s}{A_{wo} + A_{so}}$$

$$U_{zi} = \frac{\left(A_{wo} U_{zw}^{1/2} + A_{so} U_{zs}^{1/2} \right)^2}{(A_{wo} + A_{so})^2}$$

where

A_{wo} = The primary epoch year harvested winter wheat area in the pseudo zone.

A_{so} = The primary epoch year harvested spring wheat area in the pseudo zone.

U_{zw} = The squared prediction error of the winter wheat yield estimate for the pseudo zone.

U_{zs} = The squared prediction error of the spring wheat yield estimate for the pseudo zone.

B8.3 PREDICTION AT THE REGIONAL AND COUNTRY LEVELS

- a. Suppose a region consists of R zones with production estimates $P_{r1}, P_{r2}, \dots, P_{rR}$. Then the regional production prediction is obtained by

$$P_r = \sum_{z=1}^H P_{rz}$$

- b. If a country consists of K regions with production estimates $P_1, P_2, \dots, P_K$, then the country production prediction is given by

$$P = \sum_{r=1}^K P_r$$

B8.4 PREDICTION ERROR ESTIMATES FOR THE REGIONAL AND COUNTRY LEVELS.

- a. An estimate of the squared prediction error of P_r , the regional production estimate is given by

$$S_r^2 = \sum_{z=1}^R S_{rz}^2 + \sum_{z=1}^R \sum_{\substack{z'=1 \\ z \neq z'}}^R S_{rzz'}$$

where S_{rz}^2 is the estimated squared prediction error of P_{rz} , the production for z th zone and $S_{rzz'} = 0$ if the z th and z' th zones have no yield stratum in common. Otherwise,

$$S_{rzz'} = \sum_{k=1}^C A_{rik} A_{rjk} U_{rk}$$

where U_{rk} is the squared prediction error the k th yield stratum estimate commonly applicable to the area estimate A_{rz} for the z th zone and the area estimate $A_{rz'}$ for the z' th zone, and C is the number of yield strata common to both the z th and z' th zones.

- b. The estimate of the squared prediction error of P , the country production estimate is given by

$$S^2 = \sum_{r=1}^K S_r^2$$

This computation of S^2 assumes that the regional production estimates are uncorrelated. (There is no reason to doubt this assumption at the present time.)

B9.0

ZONE AND REGIONAL YIELD AND PREDICTION ERROR ESTIMATION

When there is a single yield model in a zone, the squared prediction error of the zone yield estimate provides the variance estimate needed for evaluation, i.e., computation of standard statistics. In case of more than one yield model in a zone, the average yield estimate and the associated prediction error estimate will be computed using the following formulas:

The average yield estimate for some space(s) is obtained by

$$\bar{Y}_s = \frac{P_s}{A_s}$$

where P_s is the production estimate and A_s is the area estimate for the space (Zone/region/country). An estimate of the squared prediction error of this average yield estimate is obtained by

$$U_s = \left[\bar{Y}_s^2 \frac{S_s^2}{P_s^2} + \frac{V_s}{A_s^2} - 2 \frac{\sum Y_i V_i}{P_s A_s} \right]$$

where

S_s^2 = Estimated squared prediction error of P_s , the production estimate.

V_s = Estimated variance of A_s , the area estimate.

Y_i = Yield estimate for the i th pseudo zone.

V_i = Estimated variance of the area estimated in i th pseudo zone.

APPENDIX C

HARTLEY-RAO PROCEDURE* FOR PPS SAMPLING

Let $\{Z_i\}_{i=1}^N$ be the sizes of a population of N units for which we wish to choose a sample of n with probabilities proportional to Z_i .

(a) Define $f_0 = 0$ and accumulate the Z_i so that

$$f_1 = Z_1$$

$$f_2 = Z_1 + Z_2$$

.

.

.

$$f_N = \sum_{i=1}^N Z_i$$

(b) Let $d = \frac{f_N}{n}$ and choose a random number r between 0 and

d. Define $x_k = r + (k - 1)d$, $k = 1, 2, \dots, n$

(c) Choose the i th unit where $f_{i-1} < x_k \leq f_i$ for some value k .

~~ENDING PAGE BLANK NOT FILLED~~

* Cochran, W. G., Sampling Techniques; Wiley, 1963, p. 263

APPENDIX D

SAMPLING AND ESTIMATION PROCEDURE BASED ON NATURAL STRATA

1.0 SCOPE

This document describes the sampling and aggregation requirements for the New Sampling Strategy for LACIE Phase III and beyond. These requirements include the sampling scheme as well as aggregation procedures for estimating wheat area, yield, and production and their associated prediction error estimates.

2.0 SAMPLE SELECTION PROCEDURE

The new sampling strategy to be used in LACIE Phase III and beyond utilizes:

- (1) a stratified random sampling without replacement scheme,
- (2) "universal" strata developed according to specifications oriented toward achieving homogeneity in regard to agricultural density, soil type, and climatological conditions, and
- (3) the 5x6 nautical mile segment as the sampling unit.

The total sample size allocated to the area of interest should be such that enough segments will be available for CAMS processing to produce a preassigned (expected) coefficient of variation for the estimate of production allowing for errors

ENDING PAGE BLANK NOT FILLED

due to (a) sampling, (b) classification, (c) yield prediction, and (d) loss of data.¹

3.0

ALLOCATION OF SEGMENTS

Allocation of segments refers to the determination of the total number of segments to be distributed among the defined strata. These determinations will be completed, first of all, for the universal strata (i.e., the strata developed irrespective of political subdivision level boundaries). The sample sizes determined for the universal strata will then be apportioned to the "refined" strata based upon the proportion (relative to the universal strata) of historical wheat, from the epoch year, present in the "refined" strata. The terminology "refined" strata refers to the strata that result from the intersection of the universal strata with zone-level political subdivisions in countries with detailed historical data; in countries without detailed historical data, the intersection is taken with strata²-level political subdivisions. The following sections (3.1 and 3.2) describe the allocation process to be applied to the universal strata.

¹The choice for the preassigned value for the production coefficient of variation is dependent upon the desired probabilistic accuracy of the production estimate for the area of interest. For example, if a country level production estimate is to be made, the goal is to obtain a country production estimate, at-harvest, which is within 10% of the actual production with a probability of .9. This is referred to as the 90/90 criterion.

²The use of the word "strata" in this context refers specifically to political strata; e.g., strata-level subdivisions in the USSR are oblasts.

3.1

DETERMINATION OF TOTAL SAMPLE SIZE

The total sample size allocated to the area of interest should be such that the LACIE error budget is achieved for the area estimate allowing for errors due to (a) sampling, (b) classification, (c) yield prediction and (d) loss of data. Using the best available a-priori knowledge of the magnitude of the above errors, one can approximate the mean-squared prediction error for production as a function of the total sample size, n . This equation can then be solved for n . The resulting value is given below:

$$n = \frac{\left[\sum_{j=1}^J N_j \sqrt{s_j^2 (y_j^2 + T_j^2)} \right]^2}{\delta \left[P^2 CV^2(P) - \sum_{j=1}^J A_j^2 T_j^2 \right]} \quad (3.1-1)$$

where

n = total number of segments allocated to the area of interest,

N_j = total number of ag segments in the j^{th} universal stratum,

s_j^2 = estimate of segment-to-segment variation of wheat area within the j^{th} universal stratum,

Y_j = average yield potential determined from soil characteristics in the j^{th} universal stratum,

T_j = standard deviation of yield potential in the j^{th} universal stratum,

J = total number of strata in the area of interest,

$CV(P)$ = preassigned value of the coefficient of variation of the production estimate,

$$P = \sum_{j=1}^J A_j Y_j$$

A_j = estimate of wheat area in the j^{th} universal stratum based on historical information,

and

δ = a conservative lower-bound estimate of the expected sample acquisition rate by the end of a crop season (determined from previous experience with loss of segments due to cloud cover or other reasons).

In case n is not an integer, it will be rounded upward to the integer just larger than n .

To compute the wheat area variance, S_j^2 , the following procedure will be carried out: if the j^{th} universal stratum contains at least M_j LACIE processed segments from the previous year (M_j should default to the value of 3 if no overriding value is specified), S_j^2 will be the computed segment-to-segment variation of the previously estimated LACIE wheat area. In case the number of LACIE segments is less than M_j , the estimate of the within-stratum wheat area variance will be estimated from the following regression equation:

$$S_j^2 = (\text{segment area})^2 \left[r_j^2 S_{a_j}^2 + c (a_j^2 + S_{a_j}^2) r_j (1-r_j) \right] \quad (3.1-2)$$

where

a_j = ag area in the collection of ag segments of the j^{th} universal stratum computed from a complete enumeration of ag proportion in each 5x6 nautical mile segment in the stratum as determined from a 5x6 grid overlay onto Landsat color IR imagery³,

r_j = g_j/a_j where a_j is specified above and g_j is the historical wheat area in the j^{th} universal stratum,

Sa_j^2 = the segment-to-segment within-stratum ag area variance in the collection of ag segments of the j^{th} universal stratum computed from the enumeration of the 5x6 ag proportions

and c is a constant obtained by performing a least-squares fit to other strata (each of which contains at least M_j LACIE segments-- M_j should default to the value of 3 if no overriding value is specified) in the country containing LACIE processed segments from the previous year (if too few LACIE segments are available in the country of interest, the least-squares fit should be carried out on strata from analogous areas from other countries having LACIE processed segments from the previous year).

Finally, in countries having no historical data available, since S_j^2 cannot be computed from (3.1-2) in this case, it is recommended that S_j^2 be replaced by Sa_j^2 in (3.1-1).

³SFB will produce 5x6 nautical-mile grid overlays for all areas to be sampled. These grids will be keyed to the corresponding ONC's and the percent of agricultural area will be computed for each rectangle in the grid from the Landsat imagery.

DISTRIBUTION OF SAMPLE SIZES AMONG STRATA

After determining the total number, n , of segments to be allocated to the area of interest, as indicated in the preceding section, the sample size, n_j , to allocate to the j^{th} universal stratum will be computed as follows: let t_j be the weight associated with the j^{th} universal stratum where

$$t_j = \frac{N_j \sqrt{S_j^2 (Y_j^2 + T_j^2)}}{\sum_{j=1}^J N_j \sqrt{S_j^2 (Y_j^2 + T_j^2)}} n \quad (3.2-1)$$

and where the quantities to be input into the right-hand side of (3.2-1) are as defined in section 3.1. Let n_j° be the provisional allotment to the j^{th} universal stratum where $n_j^{\circ} = \text{integer part of } t_j$. Also let

$$r_j = t_j - n_j^{\circ}$$

and

$$r = n - \sum_{j=1}^J n_j^{\circ}$$

Assign one additional sample segment to each of r of the universal strata with probabilities proportional to r_j ; i.e., let

$$n_j = n_j^0 + \delta_j, \quad j = 1, 2, \dots, J \quad (3.2-2)$$

where $\delta_j = 1$ if the j^{th} universal stratum receives an extra sample segment; $\delta_j = 0$ otherwise. The Hartley-Rao procedure⁴ will be utilized to perform the PPS allotment.

Performing the above procedure results in the determination of the universal strata sample sizes. However, wheat area and production estimates are to be made at the refined strata level and will require determinations of sample sizes for each refined stratum. For a given refined stratum, the sample size will be determined by multiplying the universal stratum sample size by the proportion (relative to the universal stratum) of historical wheat, from the epoch year, present in the given refined stratum. The fractional part of these weights account for one or more segments; consequently, these segments will be assigned to the refined strata utilizing the Hartley-Rao PPS procedure with sample sizes proportional to the fractional parts of the weights in the same manner as indicated in the previous paragraph.

3.3

DESIGNATION OF GROUPS A AND B REFINED STRATA

In the New Sampling Strategy framework, the j^{th} refined stratum is designated as a "Group A" or "Group B" refined stratum depending on whether it is allocated at least M_j or less than M_j segments, respectively, where M_j defaults to the value of 3 if no overriding value is specified. The qualitative definitions are as follows: Group A refined strata are those which represent marginal to high wheat producing areas whereas Group B refined strata represent areas having very little or no wheat historically. There is one primary exception to this

⁴c.f. Cochran, W. G., Sampling Techniques; Wiley, 1963, p. 263.

general qualification: since LACIE does not get coverage on every segment for every pass, any refined stratum having less than M_j segments after such losses shall be placed in the Group B category. If M_j or more segments subsequently become available for aggregation, the refined stratum is reinstated as Group A.

4.0

GUIDELINES FOR LOCATING SEGMENTS

The location of segments within a refined stratum will be performed by simple random sampling without replacement within the previously designated area of the stratum. All the segments with 5% or more ag within the refined stratum will be candidates for selection. After the selection of the sample, the actual acquisitions will be located on a mosaic. The latitude and longitude of the center of each segment is obtained and delivered to DAPTS for ordering/back-ordering.

5.0

APPORTIONMENT PROCEDURES

5.1

APPORTIONMENT OF POLITICAL SUBDIVISION DATA TO STRATA

As noted in section 3.1, it may be necessary to know the universal stratum historical wheat area g_j in order to compute r_j for input into (3.1-2). This quantity will also be needed for each refined stratum in the aggregation. In countries having historical data available only on one level smaller than the country itself (e.g., the oblast level in the U.S.S.R.), an apportionment of the political subdivision wheat proportional to the universal/refined stratum must be performed. This will be accomplished as follows:

$$g_j = \frac{a_{ij}}{a_i} w_i \quad (5.1-1)$$

where

g_j = the apportioned estimate of historical wheat area in the j^{th} universal/refined stratum,

a_{ij} = the ag area in the collection of ag segments common to the j^{th} universal/refined stratum and i^{th} political subdivision,

w_i = the wheat area in the i^{th} political subdivision based on historical data,

and

a_i = the total ag area in the collection of ag segments of the i^{th} political subdivision.

5.2 APPORTIONMENT OF AREA ESTIMATES TO YIELD STRATA

The Sampling Strategy requires that yield and area (universal) strata be coincident. The situation may arise, however, where this may not be the case. In such cases an apportionment of the acreage estimate to the yield strata is needed to permit the estimate of production at political subdivision (or other) levels. The following discussion specifies the requirements in support of the New Sampling Strategy in this case.

Another set of area strata will be generated consisting of the collection of all the area "sub-strata" that result from the intersection of the refined strata with the yield strata. This will result in each yield stratum containing one or more area "substrata." The LACIE wheat area estimate for each area "substratum" will be obtained by apportioning out the LACIE wheat area estimate of the refined stratum according to the proportion of ag area in each "substratum" relative to the refined stratum from which it came.

Referring to Illustration 1 below, if the indicated yield strata covered the area of interest with $\{A_1, A_2, A_3, A_4, A_5\}$ being the collection of refined strata over this same area, then, in view of the lack of coincidence of area and yield strata, a new collection of area substrata is generated consisting of $\{A_{11}, A_{12}, A_{21}, A_{22}, A_{31}, A_{32}, A_{41}, A_{51}, A_{52}, A_{53}, A_{54}\}$.

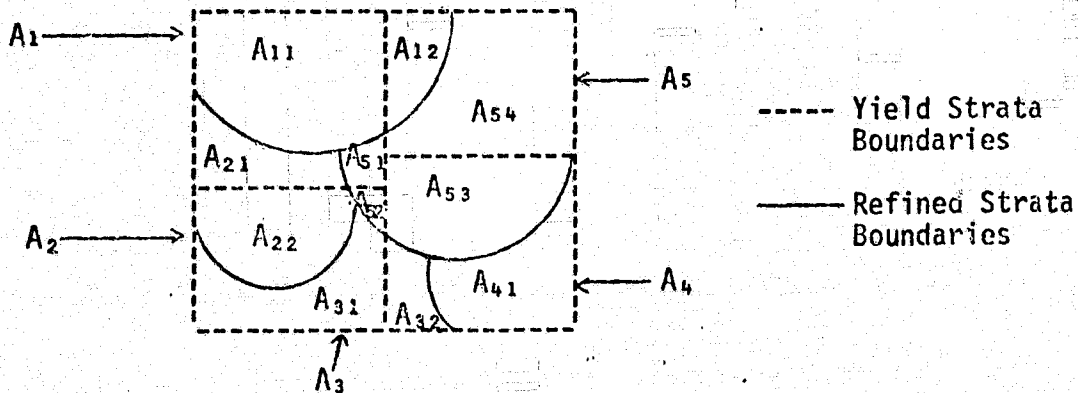


ILLUSTRATION 1: A Refinement of Acreage Strata to the Yield Strata

Referring to the above illustration, if $\hat{W}_{ij}$ is used to denote the LACIE estimated wheat area apportioned to A_{ij} , then

$$\hat{W}_{ij} = \begin{cases} \frac{a_{ji}}{a_j} A_{jA}, & \text{if the } i^{\text{th}} \text{ substratum is contained} \\ & \text{within the } j^{\text{th}} \text{ Group A (see section} \\ & \text{6.1 for } A_{jA}) \text{ refined stratum,} \\ \\ \frac{a_{ji}}{a_j} A_{jB}, & \text{if the } i^{\text{th}} \text{ substratum is contained} \\ & \text{within the } j^{\text{th}} \text{ Group B (see section} \\ & \text{6.2 for } A_{jB}) \text{ refined stratum,} \end{cases} \quad (5.2-1)$$

where

a_{ji} = the total ag area of A_{ji} determined from the product of the ag proportion A_{ji} (determined from the complete enumeration of ag in the 5x6's) with the planimetered area of A_{ji} ,

and

a_j = the total ag area of A_j (determined from the complete enumeration of ag in the 5x6's).

The above procedure will identify the area "substrata" and provide associated wheat area estimates to be aggregated within each yield stratum to compute production. It should be observed that the procedures discussed in this section will necessarily be needed for use in the estimate of production and production prediction error in case the yield strata are not coincident with the universal strata; however, the wheat area estimate for a given area of interest will be the same regardless of whether the area aggregation is performed from the refined strata or from the substrata (generated from intersecting the refined strata with the yield strata).

6.0 STRATUM AREA ESTIMATION

6.1 GROUP A REFINED STRATUM AREA ESTIMATE

The wheat area estimate for the j^{th} Group A refined stratum is calculated as follows:

$$A_{jA} = \frac{N_{jA}}{M_{jA}} \sum_{k=1}^{M_{jA}} A_{jKA} \quad (6.1-1)$$

where

- A_{jA} = the LACIE wheat area estimate for the j^{th} Group A refined stratum,
- N_{jA} = the total number of ag segments in the j^{th} Group A refined stratum,
- M_{jA} = the total number of sample segments for which classification data is available in the j^{th} Group A refined stratum,
- A_{jka} = the LACIE wheat area estimate for the k^{th} sample segment in the j^{th} Group A refined stratum.

6.2

GROUP B REFINED STRATUM AREA ESTIMATE

The wheat area estimate of the j^{th} Group B acreage stratum is calculated as follows:

$$A_{jB} = \sum_{k=1}^m C_{kj} A_{kA} + \left(n_{jB} / M_j \right) A'_{jB} \quad (6.2-1)$$

where

$$C_{kj} = \left(1 - \frac{n_{jB}}{M_j} \right) \left(\frac{d_{kj} W_{jB}}{\sum_{k=1}^m d'_{kj} W'_{kA}} \right)$$

W_{jB} = the most recent epoch year (or average over previous two or three years) harvested wheat area in the j^{th} Group B stratum.

$$d_{kj} = \begin{cases} 1, & \text{if the } k^{\text{th}} \text{ Group A stratum is} \\ & \text{used in the estimate of the} \\ & j^{\text{th}} \text{ Group B stratum} \\ 0, & \text{otherwise} \end{cases}$$

$$\sum_{k=1}^m d_{kj} A_{kA} = \text{the sum of the LACIE estimates of wheat area in all Group A strata that are used to estimate the given } j^{\text{th}} \text{ Group B stratum,}$$

$$\sum_{k=1}^m d'_{kj} W'_{kA} = \text{the sum of the most recent epoch year (or average over previous two to three years) harvested wheat area in the Group A strata that are used to estimate the } j^{\text{th}} \text{ Group B stratum,}$$

m = the total number of Group A strata in the area of interest plus any additional Group A strata that are not contained within the area of interest but that are used in the ratio estimation of the Group B strata contained in the area of interest,

n_{jB} = the total number of sample segments for which classification data is available in the j^{th} Group B stratum,

and

$A'_{jB} = 0$ if $n_{jB} = 0$; otherwise

$$A'_{jB} = \frac{N_{jB}}{n_{jB}} \sum_{k=1}^{n_{jB}} A_{jkB} \quad (6.2-2)$$

where

N_{jB} = the total number of ag segments in the j^{th} Group B refined stratum

A_{jkB} = the LACIE wheat area estimate for the k^{th} sample segment in the j^{th} Group B stratum.

and the capability should exist for M_j to default to the value of 3 if no overriding value is specified. (Recall that M_j is the minimal number of segments that must be acquired in the j^{th} refined stratum before it is designated as a Group A stratum.)

It should be re-emphasized at this point that the most refined level for which historical wheat area is available in many LACIE foreign areas is one level below the country level (Crop Region or Economic Region). Consequently, the historical wheat values for each stratum utilized in the Group B estimator must be apportioned to the stratum as indicated in section 5.1.

In the preceding discussion, the Group A strata to be utilized as a base for the ratio-estimated part of the wheat area estimate of a given Group B stratum are to be selected according to the following guidelines:

- (1) First of all, the capability should exist to permit interactive selection of the appropriate Group A strata as a base in the ratio-estimated part of a given Group B stratum.
- (2) If (1) is not available (or, if available, but not utilized) the software should default to the use of all Group A strata (if there is at least one) in the zone (containing the given Group B stratum).

7.0

STRATUM WHEAT AREA VARIANCE

7.1

GROUP A REFINED STRATUM VARIANCE

The variance of the estimate of wheat area for the j^{th} Group A refined stratum is given by:

$$V_{jA} = \frac{N_{jA}^2}{M_{jA}} S_{jA}^2 \quad (7.1-1)$$

where

V_{jA} = the variance of the j^{th} Group A refined stratum estimate of wheat area,

$$S_{jA}^2 = 1 / (M_{jA} - 1) \sum_{k=1}^{M_{jA}} (A_{jkA} - \bar{A}_{jA})^2 \quad (7.1-2)$$

and

$$\bar{A}_{jA} = (1 / M_{jA}) \sum_{k=1}^{M_{jA}} A_{jkA} \quad (7.1-3)$$

7.2

GROUP B STRATUM VARIANCE

The variance of the estimate of wheat area for the j^{th} Group B stratum is given by:

$$V_{jB} = \sum_{k=1}^m C_{kj}^2 V_{kA} + (n_{jB} / M)^2 V'_{jB} \quad (7.2-1)$$

where V'_{jB} is the estimate of segment-to-segment variance of wheat area computed for the refined stratum in the same manner as indicated in section 3.1 (i.e., by making use of the allocation data at the refined stratum level), if $n_{jB} \leq 1$; otherwise (i.e., if $2 \leq n_{jB} < M_j$, V'_{jB} is estimated directly (i.e., in the same manner as the estimate of V_{jA} -- see equation (7.1-1)).

8.0 HIGHER LEVEL ESTIMATES

8.1 AREA AGGREGATION TO THE ZONE, REGION, AND COUNTY LEVELS

The wheat area estimate, A , of the area of interest (whether it be zone, region, or country) is given by:

$$A = \sum_{j=1}^m n_j A_{jA} + \sum_{i=1}^b A_{iB} \quad (8.1-1)$$

where

$$n_j = \begin{cases} 1, & \text{if } j^{\text{th}} \text{ Group A refined stratum is} \\ & \text{contained within the area of interest} \\ 0, & \text{otherwise} \end{cases}$$

and m is the total number of Group A strata that are contained in the area of interest plus any additional Group A strata that are not contained within the area of interest but that are used in the ratio estimation of the Group B strata contained in the area of interest. The quantity b denotes the total number of Group B strata in the area of interest.

8.2

VARIANCE AGGREGATION TO THE ZONE, REGION, AND COUNTRY LEVELS

After substituting the expression in equation (6.2-1) into (8.1-1) and simplifying, it is straightforward to see that A in (8.1-1) is expressible as:

$$A = \sum_{j=1}^m \alpha_j A_{jA} + \sum_{i=1}^b \left(n_{iB} A'_{iB} / M_i \right) . \quad (8.2-1)$$

where

$$\alpha_j = n_j + \sum_{i=1}^b c_{ji}$$

Consequently, the variance, V_A , for the estimate of wheat area for the area of interest is given by:

$$V_A = \sum_{j=1}^m \alpha_j^2 V_{jA} + \sum_{i=1}^b \left(n_{iB} / M_i \right) V'_{jB} \quad (8.2-2)$$

9.0

SPRING AND WINTER WHEAT

9.1

AREA AND VARIANCE ESTIMATION

In a mixed wheat area, separate area estimates will be made for the winter and the spring wheat using the aggregation procedures described in sections 6.0 through 8.2 with inputs (CAMS/CAS as well as historical) of winter wheat for a winter wheat aggregation and those of spring wheat for a spring

wheat aggregation. This will provide the spring wheat and winter wheat area estimates and their respective variance estimates at the strata and zone levels. The procedures in sections 8.1 and 8.2 are used to obtain separate winter wheat and spring wheat area estimates and the corresponding variance estimates at the zone, regional, and country levels.

9.2

TOTAL WHEAT: AREA AND VARIANCE ESTIMATION

The total wheat area estimate in a mixed wheat area will be computed by aggregating the winter wheat and the spring wheat area estimates for the area, that is, if A_w and A_s denote the winter and spring wheat area estimates, respectively, the total wheat area estimate, A_t , is given by

$$A_t = A_w + A_s. \quad (9.2-1)$$

This will be done at the zone and higher levels.

The variance estimates for the total wheat at the zone and higher levels will be those obtained from the total wheat aggregation to be made with inputs of total wheat by CAS for the segments and historical data. The procedure is the same as described in section 8.2.

10.0

STANDARD OUTPUT STATISTICS

The standard statistics for area estimation and aggregation are composed of standard deviation (σ), coefficient of variation (CV); 90 percent confidence limits (CL), and probability of 10 percent relative error (Ω). If A_s and V_s are the area and variance estimates at some level s , statistics will be calculated as follows:

$$a. \sigma_s = \sqrt{V_s} \quad (10.0-1)$$

$$b. CV_s = \frac{\sigma_s}{A_s} \quad (10.0-2)$$

$$c. CL_s^{(+)} = A_s + (1.645)\sigma_s \quad (10.0-3a)$$

$$CL_s^{(-)} = A_s - (1.645)\sigma_s \quad (10.0-3b)$$

$$d. W_s = 2 p(x) - 1 \quad (10.0-4)$$

$$\text{where } p(x) = 1 - z(x) (a_1 t + a_2 t^2 + a_3 t^3)$$

$$\text{with } t = \frac{1}{1 + px}$$

$$x = \frac{.10}{CV_s}$$

$$p = .3327$$

$$a_1 = .4362$$

$$a_2 = -.1202$$

$$a_3 = .9373$$

and

$$z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

11.0 PRODUCTION ESTIMATION

11.1 PRODUCTION AND VARIANCE ESTIMATION AT STRATUM LEVEL

The estimate of wheat production for the j^{th} area refined stratum (stratum as used here does not refer to CRD, oblasts, etc.) is given by:

$$P_j = \begin{cases} A_{jA} Y_{jA} & , \text{ if the } j^{\text{th}} \text{ refined stratum is} \\ & \text{a Group A stratum} \\ A_{jB} Y_{jB} & , \text{ if the } j^{\text{th}} \text{ refined stratum is} \\ & \text{a Group B stratum} \end{cases} \quad (11.1-1)$$

In the above, Y_{jA} (Y_{jB}) is the predicted yield for the j^{th} Group A (Group B) refined stratum as given by the yield estimation model. The estimate of the prediction error of the j^{th} Group A refined stratum production estimate is given by:

$$z_{jA}^2 = V_{jA} Y_{jA}^2 + T_{jA}^2 A_{jA}^2 - V_{jA} T_{jA}^2 \quad (11.1-2)$$

where A_{jA} and Y_{jA} are the estimates of wheat area and yield, respectively, of the j^{th} Group A refined stratum and

V_{jA} = estimate of the wheat area variance for the j^{th} Group A refined stratum,

and

T_{jA}^2 = the estimated squared prediction error of yield in the j^{th} Group A refined stratum.

The prediction error of the j^{th} Group B refined stratum production estimate is similarly obtained.

In case the yield strata are not coincident with the universal strata, the estimate of production and its associated prediction error estimate at the substratum level is given below. In particular, if we let P_{ijr} denote the production estimate of the i^{th} substratum (apportioned from the j^{th} yield stratum which has yield Y_r , then

$$P_{ijr} = \hat{W}_{ij} Y_r \quad (11.1-3)$$

where $\hat{W}_{ij}$ is given by equation (5.2-1). Moreover, denoting the production prediction error estimate of P_{ijr} by z_{ijr}^2 , we have

$$z_{ijr}^2 = v_{ij} y_r^2 + T_r^2 \hat{W}_{ij}^2 - v_{ij} T_r^2 \quad (11.1-4)$$

where $\hat{W}_{ij}$ and Y_r are the estimates of wheat area and yield, respectively, of the i^{th} substratum (contained in the j^{th} refined stratum), and

$$v_{ij} = \begin{cases} \left(\frac{a_{ji}}{a_j} \right)^2 v_{jA} , & \text{if the } i^{\text{th}} \text{ substratum is} \\ & \text{contained within the } j^{\text{th}} \\ & \text{Group A refined stratum,} \\ \left(\frac{a_{ji}}{a_j} \right)^2 v_{jB} , & \text{if the } i^{\text{th}} \text{ substratum is} \\ & \text{contained within the } j^{\text{th}} \\ & \text{Group B refined stratum.} \end{cases} \quad (11.1-5)$$

The quantities a_{ji} and a_j are as defined in section 5.2.

11.2

PRODUCTION ESTIMATE FOR A ZONE, REGION, OR COUNTRY LEVEL

The wheat production estimate, P_A , for the area of interest (whether it be zone, region, or country level) is given by equation (11.2-1) or equation (11.2-2), respectively, depending on whether or not the yield strata coincide with the universal strata, i.e.

$$P_A = \sum_{r=1}^H \left[\sum_{j=1}^m \gamma_{rj} n_j A_{jA} + \sum_{i=1}^b \beta_{ri} A_{iB} \right] y_r \quad (11.2-1)$$

or

$$P_A = \sum_{r=1}^H \left[\sum_{j=1}^m \gamma_{rj} n_j q_{jA} A_{jA} + \sum_{i=1}^b \beta_{ri} q_{iB} A_{iB} \right] y_r \quad (11.2-2)$$

That is,

$$P_A = \sum_{r=1}^H A_r y_r$$

where

$$A_r = \sum_{j=1}^m \left[\gamma_{rj} n_j + \sum_{i=1}^b \beta_{ri} C_{ji} \right] A_{jA} + \sum_{i=1}^b \left(\beta_{ri} n_{iB} A'_{iB} / M_i \right) \quad (11.2-3)$$

or A_r is given by this same expression after replacing $\gamma_{rj} n_j$ by $\gamma_{rj} n_j q_{jA}$ and β_{ri} by $\beta_{ri} q_{iB}$. In the situation where the universal strata and yield strata coincide, equation (11.2-1) applies where

Y_r = the LACIE predicted yield in the r^{th} yield stratum,

H = the total number of yield strata in the area of interest,

$$\gamma_{rj} = \begin{cases} 1, & \text{if the } j^{th} \text{ Group A stratum lies} \\ & \text{within (or coincides with) the } r^{th} \\ & \text{yield stratum} \\ 0, & \text{otherwise} \end{cases}$$

$$\beta_{ri} = \begin{cases} 1, & \text{if the } i^{th} \text{ Group B stratum lies} \\ & \text{within (or coincides with) the } r^{th} \\ & \text{yield stratum} \\ 0, & \text{otherwise} \end{cases}$$

and M_i is a variable that defaults to the value of 3 if no overriding value is specified.

In the situation where the universal strata and yield strata do not coincide, equation (11.2-2) applies with Y_r and H as defined above; however,

$$\gamma_{rj} = \begin{cases} 1, & \text{if the } j^{th} \text{ Group A stratum intersects} \\ & \text{the } r^{th} \text{ yield stratum} \\ 0, & \text{otherwise} \end{cases}$$

$$\beta_{ri} = \begin{cases} 1, & \text{if the } i^{th} \text{ Group B stratum intersects} \\ & \text{the } r^{th} \text{ yield stratum} \\ 0, & \text{otherwise,} \end{cases}$$

$$q_{jA} = \sum_{q=1}^{m_j} (a_{jq}/a_j) \quad (11.2-4)$$

and

$$q_{iB} = \sum_{q=1}^{b_i} (a_{iq}/a_i) \quad (11.2-5)$$

In equations (11.2-4) and 11.2-5), the small a's are as given in section 5.2, m_j is the total number of substrata apportioned out of the j^{th} refined Group A stratum and that lies within the r^{th} yield stratum. Similarly, b_i is the total number of substrata apportioned out of the i^{th} refined Group B stratum and that lies within the r^{th} yield stratum.

11.3

PRODUCTION PREDICTION ERROR FOR A ZONE, REGION, OR COUNTRY LEVEL

The estimate of the squared prediction error, S_p^2 , of the production estimate for the area of interest (whether it be zone, region, or country level) is given by

$$S_p^2 = \sum_{r=1}^H \left[Y_r^2 V_{rA} + A_r^2 T_r^2 - V_{rA} T_r^2 \right] + 2 \sum_{r=2}^H \sum_{r'=1}^{r-1} Y_r Y_{r'} \text{COV} (A_r, A_{r'}) \quad (11.3-1)$$

and

$$\text{COV} (A_r, A_{r'}) = \sum_j \alpha_{jr} \alpha_{jr'} V_{jA}$$

where

$$\alpha_{jr} = \gamma_{rj} n_j + \sum_{i=1}^b \beta_{ri} C_{ji} ,$$

T_r^2 = the estimated squared prediction error of yield in the r^{th} yield stratum,

and

$$\begin{aligned} V_{rA} = & \sum_{j=1}^m \left[\gamma_{rj} n_j + \sum_{i=1}^b \beta_{ri} C_{ji} \right]^2 V_{jA} \\ & + \sum_{i=1}^b \left(\beta_{ri} n_{iB} / M_i \right)^2 V_{iB} \end{aligned} \quad (11.3-2)$$

The preceding equations apply, of course, to the situation where the universal strata and yield strata are coincident. If this should not be the case, the same equations would still apply provided $\gamma_{rj} n_j$ and β_{ri} are replaced with $\gamma_{rj} n_j q_{jA}$ and $\beta_{ri} q_{iB}$, respectively, throughout.

12.0

SPRING AND WINTER WHEAT PRODUCTION

12.1

PRODUCTION ESTIMATION IN MIXED WHEAT AREAS

In a mixed wheat area, separate production and predicted production error estimates will be made for the winter wheat and spring wheat using the procedures described in section 11.1 - 11.3. The total wheat production estimate in a mixed wheat area will be computed by aggregating the winter production and the spring production, that is, if P_w and P_s denote the winter and spring wheat production

estimates, respectively, the total wheat production estimate, P_t , is given by:

$$P_t = P_w + P_s \quad (12.1-1)$$

12.2 PRODUCTION ERROR ESTIMATES IN MIXED WHEAT AREAS

The estimate of the production prediction error at the area of interest level (whether it be zone, region, or country level) in a mixed wheat area is given by equations (11.3-1) - (11.3-3) with the following modifications: if the j^{th} Group A re-fined stratum contains mixed wheat and is supplied with both a spring, y_r^s , and winter, y_r^w , yield estimate then y_r and T_r^2 in (11.3-1) and 11.3-2) are to be replaced by:

$$y_r = \frac{A_{wr}y_r^w + A_{sr}y_r^s}{A_{wr} + A_{sr}} \quad (12.2-1)$$

and

$$T_r^2 = \left(\frac{A_{wr}T_r^w + A_{sr}T_r^s}{A_{wr} + A_{sr}} \right)^2, \quad (12.2-2)$$

respectively, where

A_{wr} = the epoch year harvested winter wheat area in the r^{th} yield stratum,

A_{sr} = the epoch year harvested spring wheat area in the r^{th} yield stratum,

T_r^W = the root mean square of the prediction error of the winter wheat yield estimate for the r^{th} yield stratum,

and

T_r^S = the root mean square of the prediction error of the spring wheat yield estimate for the r^{th} yield stratum.

13.0

AREA OF INTEREST YIELD AND PREDICTION ERROR ESTIMATION

When there is a single yield model covering the area of interest, the squared prediction error of the yield model provides the variance needed for evaluation, i.e., for computation of standard statistics. In case of more than one yield model in the area of interest (whether it be zone, region, or country), the average yield estimate and the associated prediction error estimate will be computed using the following formulas:

The average yield estimate for the area, C, of interest is obtained by:

$$\bar{Y}_C = \frac{P_C}{A_C}$$

where P_C is the production estimate and A_C is the area estimate for the area of interest. An estimate of the squared prediction error of this average yield estimate is obtained by:

$$U_C = \bar{Y}_C^2 \left[\frac{s_C^2}{P_C^2} + \frac{V_C}{A_C^2} - \frac{2 \sum_{i=1}^H Y_i V_i}{P_C A_C} \right] \quad (13.0-1)$$

where

S_C^2 = estimated squared prediction error of P_C ,
the production estimate,

V_C = estimated prediction error of A_C , the area
estimate,

Y_i = yield estimate for the i^{th} yield stratum,

and

V_i = estimated prediction error of the area
estimate in the i^{th} yield stratum (com-
puted from use of equation (8.2-2) where
the area of interest referred to by that
estimate is the i^{th} yield stratum in this
case).